

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**



## NATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

International Patent Classification 6:

C07K 14/00, A01N 37/18, 43/04, C07H  
17/00, C12N 5/00, C12P 21/06

A1

(11) International Publication Number:

WO 96/27610

(43) International Publication Date: 12 September 1996 (12.09.96)

(21) International Application Number: PCT/US96/03172

(22) International Filing Date: 7 March 1996 (07.03.96)

(30) Priority Data: 400,159 7 March 1995 (07.03.95) US

(60) Parent Application or Grant

(63) Related by Continuation

US 400,159 (CIP)  
Filed on 7 March 1995 (07.03.95)

(71) Applicants (for all designated States except US): YALE UNIVERSITY [US/US]; 451 College Street, New Haven, CT 06520 (US). IMPERIAL CANCER RESEARCH TECHNOLOGY, LTD. [GB/GB]; Sardinia House, Sardinia Street, London WC2A 3NL (GB).

(72) Inventors; and

(75) Inventors/Applicants (for US only): ISH-HOROWICZ, David [GB/GB]; 2 Lower Fisher Row, Oxford OX1 1JY (GB). HENRIQUE, Domingos, M., P. [PT/GB]; 1A Summerhill Road, Oxford OX2 7JY (GB). LEWIS, Julian, H. [GB/GB]; 60 Sandfield Road, Oxford OX3 7RJ (GB). MYAT, Anna, M. [GB/GB]; 4 Chawley Lane, Cumnor Hill, Oxford OX2

9PX (GB). ARTAVANIS-TSAKONAS, Spyridon [US/US]; 192 Ridgewood Avenue, Hamden, CT 06517 (US). MANN, Robert, S. [US/US]; 21 Towne House Road, Hamden, CT 06514 (US). GRAY, Grace, E. [US/US]; Apartment 4, 815 Whitney Avenue, New Haven, CT 06511 (US).

(74) Agents: MISROCK, S., Leslie et al.; Pennie &amp; Edmonds, 1155 Avenue of the Americas, New York, NY 10036 (US).

(81) Designated States: AL, AM, AU, AZ, BB, BG, BR, BY, CA, CN, CZ, EE, FI, GE, HU, IS, JP, KG, KP, KR, KZ, LK, LR, LS, LT, LV, MD, MG, MK, MN, MX, NO, NZ, PL, RO, RU, SG, SI, SK, TJ, TM, TR, TT, UA, US, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: NUCLEOTIDE AND PROTEIN SEQUENCES OF VERTEBRATE SERRATE GENES AND METHODS BASED THEREON

10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210 220 230 240 250 260 270 280 290 300 310 320 330 340 350 360 370 380 390 400 410 420 430 440 450 460 470 480 490 500 510 520 530 540 550 560 570 580 590 600 610 620 630 640 650 660 670 680 690 700 710 720 730 740 750 760 770 780 790 800 810 820 830 840 850 860 870 880 890 900 910 920 930 940 950 960 970 980 990 1000 1010 1020 1030 1040 1050 1060 1070 1080 1090 1100 1110 1120 1130 1140 1150 1160 1170 1180 1190 1200 1210 1220 1230 1240 1250 1260 1270 1280 1290 1300 1310 1320 1330 1340 1350 1360 1370 1380 1390 1400 1410 1420 1430 1440 1450 1460 1470 1480 1490 1500 1510 1520 1530 1540 1550 1560 1570 1580 1590 1600 1610 1620 1630 1640 1650 1660 1670 1680 1690 1700 1710 1720 1730 1740 1750 1760 1770 1780 1790 1800 1810 1820 1830 1840 1850 1860 1870 1880 1890 1900 1910 1920 1930 1940 1950 1960 1970 1980 1990 2000 2010 2020 2030 2040 2050 2060 2070 2080 2090 2100 2110 2120 2130 2140 2150 2160 2170 2180 2190 2200 2210 2220 2230 2240 2250 2260 2270 2280 2290 2300 2310 2320 2330 2340 2350 2360 2370 2380 2390 2400 2410 2420 2430 2440 2450 2460 2470 2480 2490 2500 2510 2520 2530 2540 2550 2560 2570 2580 2590 2600 2610 2620 2630 2640 2650 2660 2670 2680 2690 2700 2710 2720 2730 2740 2750 2760 2770 2780 2790 2800 2810 2820 2830 2840 2850 2860 2870 2880 2890 2900 2910 2920 2930 2940 2950 2960 2970 2980 2990 3000 3010 3020 3030 3040 3050 3060 3070 3080 3090 3100 3110 3120 3130 3140 3150 3160 3170 3180 3190 3200 3210 3220 3230 3240 3250 3260 3270 3280 3290 3300 3310 3320 3330 3340 3350 3360 3370 3380 3390 3400 3410 3420 3430 3440 3450 3460 3470 3480 3490 3500 3510 3520 3530 3540 3550 3560 3570 3580 3590 3600 3610 3620 3630 3640 3650 3660 3670 3680 3690 3700 3710 3720 3730 3740 3750 3760 3770 3780 3790 3800 3810 3820 3830 3840 3850 3860 3870 3880 3890 3900 3910 3920 3930 3940 3950 3960 3970 3980 3990 4000 4010 4020 4030 4040 4050 4060 4070 4080 4090 4100 4110 4120 4130 4140 4150 4160 4170 4180 4190 4200 4210 4220 4230 4240 4250 4260 4270 4280 4290 4300 4310 4320 4330 4340 4350 4360 4370 4380 4390 4400 4410 4420 4430 4440 4450 4460 4470 4480 4490 4500 4510 4520 4530 4540 4550 4560 4570 4580 4590 4600 4610 4620 4630 4640 4650 4660 4670 4680 4690 4700 4710 4720 4730 4740 4750 4760 4770 4780 4790 4800 4810 4820 4830 4840 4850 4860 4870 4880 4890 4900 4910 4920 4930 4940 4950 4960 4970 4980 4990 5000 5010 5020 5030 5040 5050 5060 5070 5080 5090 5100 5110 5120 5130 5140 5150 5160 5170 5180 5190 5200 5210 5220 5230 5240 5250 5260 5270 5280 5290 5300 5310 5320 5330 5340 5350 5360 5370 5380 5390 5400 5410 5420 5430 5440 5450 5460 5470 5480 5490 5500 5510 5520 5530 5540 5550 5560 5570 5580 5590 5600 5610 5620 5630 5640 5650 5660 5670 5680 5690 5700 5710 5720 5730 5740 5750 5760 5770 5780 5790 5800 5810 5820 5830 5840 5850 5860 5870 5880 5890 5900 5910 5920 5930 5940 5950 5960 5970 5980 5990 6000 6010 6020 6030 6040 6050 6060 6070 6080 6090 6100 6110 6120 6130 6140 6150 6160 6170 6180 6190 6200 6210 6220 6230 6240 6250 6260 6270 6280 6290 6300 6310 6320 6330 6340 6350 6360 6370 6380 6390 6400 6410 6420 6430 6440 6450 6460 6470 6480 6490 6500 6510 6520 6530 6540 6550 6560 6570 6580 6590 6600 6610 6620 6630 6640 6650 6660 6670 6680 6690 6700 6710 6720 6730 6740 6750 6760 6770 6780 6790 6800 6810 6820 6830 6840 6850 6860 6870 6880 6890 6900 6910 6920 6930 6940 6950 6960 6970 6980 6990 7000 7010 7020 7030 7040 7050 7060 7070 7080 7090 7100 7110 7120 7130 7140 7150 7160 7170 7180 7190 7200 7210 7220 7230 7240 7250 7260 7270 7280 7290 7300 7310 7320 7330 7340 7350 7360 7370 7380 7390 7400 7410 7420 7430 7440 7450 7460 7470 7480 7490 7500 7510 7520 7530 7540 7550 7560 7570 7580 7590 7600 7610 7620 7630 7640 7650 7660 7670 7680 7690 7700 7710 7720 7730 7740 7750 7760 7770 7780 7790 7800 7810 7820 7830 7840 7850 7860 7870 7880 7890 7900 7910 7920 7930 7940 7950 7960 7970 7980 7990 8000 8010 8020 8030 8040 8050 8060 8070 8080 8090 8100 8110 8120 8130 8140 8150 8160 8170 8180 8190 8200 8210 8220 8230 8240 8250 8260 8270 8280 8290 8300 8310 8320 8330 8340 8350 8360 8370 8380 8390 8400 8410 8420 8430 8440 8450 8460 8470 8480 8490 8500 8510 8520 8530 8540 8550 8560 8570 8580 8590 8600 8610 8620 8630 8640 8650 8660 8670 8680 8690 8700 8710 8720 8730 8740 8750 8760 8770 8780 8790 8800 8810 8820 8830 8840 8850 8860 8870 8880 8890 8900 8910 8920 8930 8940 8950 8960 8970 8980 8990 9000 9010 9020 9030 9040 9050 9060 9070 9080 9090 9100 9110 9120 9130 9140 9150 9160 9170 9180 9190 9200 9210 9220 9230 9240 9250 9260 9270 9280 9290 9300 9310 9320 9330 9340 9350 9360 9370 9380 9390 9400 9410 9420 9430 9440 9450 9460 9470 9480 9490 9500 9510 9520 9530 9540 9550 9560 9570 9580 9590 9600 9610 9620 9630 9640 9650 9660 9670 9680 9690 9700 9710 9720 9730 9740 9750 9760 9770 9780 9790 9800 9810 9820 9830 9840 9850 9860 9870 9880 9890 9900 9910 9920 9930 9940 9950 9960 9970 9980 9990 10000 10010 10020 10030 10040 10050 10060 10070 10080 10090 10100 10110 10120 10130 10140 10150 10160 10170 10180 10190 10200 10210 10220 10230 10240 10250 10260 10270 10280 10290 10300 10310 10320 10330 10340 10350 10360 10370 10380 10390 10400 10410 10420 10430 10440 10450 10460 10470 10480 10490 10500 10510 10520 10530 10540 10550 10560 10570 10580 10590 10600 10610 10620 10630 10640 10650 10660 10670 10680 10690 10700 10710 10720 10730 10740 10750 10760 10770 10780 10790 10800 10810 10820 10830 10840 10850 10860 10870 10880 10890 10900 10910 10920 10930 10940 10950 10960 10970 10980 10990 11000 11010 11020 11030 11040 11050 11060 11070 11080 11090 11100 11110 11120 11130 11140 11150 11160 11170 11180 11190 11200 11210 11220 11230 11240 11250 11260 11270 11280 11290 11300 11310 11320 11330 11340 11350 11360 11370 11380 11390 11400 11410 11420 11430 11440 11450 11460 11470 11480 11490 11500 11510 11520 11530 11540 11550 11560 11570 11580 11590 11600 11610 11620 11630 11640 11650 11660 11670 11680 11690 11700 11710 11720 11730 11740 11750 11760 11770 11780 11790 11800 11810 11820 11830 11840 11850 11860 11870 11880 11890 11900 11910 11920 11930 11940 11950 11960 11970 11980 11990 12000 12010 12020 12030 12040 12050 12060 12070 12080 12090 12100 12110 12120 12130 12140 12150 12160 12170 12180 12190 12200 12210 12220 12230 12240 12250 12260 12270 12280 12290 12300 12310 12320 12330 12340 12350 12360 12370 12380 12390 12400 12410 12420 12430 12440 12450 12460 12470 12480 12490 12500 12510 12520 12530 12540 12550 12560 12570 12580 12590 12600 12610 12620 12630 12640 12650 12660 12670 12680 12690 12700 12710 12720 12730 12740 12750 12760 12770 12780 12790 12800 12810 12820 12830 12840 12850 12860 12870 12880 12890 12900 12910 12920 12930 12940 12950 12960 12970 12980 12990 13000 13010 13020 13030 13040 13050 13060 13070 13080 13090 13100 13110 13120 13130 13140 13150 13160 13170 13180 13190 13200 13210 13220 13230 13240 13250 13260 13270 13280 13290 13300 13310 13320 13330 13340 13350 13360 13370 13380 13390 13400 13410 13420 13430 13440 13450 13460 13470 13480 13490 13500 13510 13520 13530 13540 13550 13560 13570 13580 13590 13600 13610 13620 13630 13640 13650 13660 13670 13680 13690 13700 13710 13720 13730 13740 13750 13760 13770 13780 13790 13800 13810 13820 13830 13840 13850 13860 13870 13880 13890 13900 13910 13920 13930 13940 13950 13960 13970 13980 13990 14000 14010 14020 14030 14040 14050 14060 14070 14080 14090 14100 14110 14120 14130 14140 14150 14160 14170 14180 14190 14200 14210 14220 14230 14240 14250 14260 14270 14280 14290 14300 14310 14320 14330 14340 14350 14360 14370 14380 14390 14400 14410 14420 14430 14440 14450 14460 14470 14480 14490 14500 14510 14520 14530 14540 14550 14560 14570 14580 14590 14600 14610 14620 14630 14640 14650 14660 14670 14680 14690 14700 14710 14720 14730 14740 14750 14760 14770 14780 14790 14800 14810 14820 14830 14840 14850 14860 14870 14880 14890 14900 14910 14920 14930 14940 14950 14960 14970 14980 14990 15000 15010 15020 15030 15040 15050 15060 15070 15080 15090 15100 15110 15120 15130 15140 15150 15160 15170 15180 15190 15200 15210 15220 15230 15240 15250 15260 15270 15280 15290 15300 15310 15320 15330 15340 15350 15360 15370 15380 15390 15400 15410 15420 15430 15440 15450 15460 15470 15480 15490 15500 15510 15520 15530 15540 15550 15560 15570 15580 15590 15600 15610 15620 15630 15640 15650 15660 15670 15680 15690 15700 15710 15720 15730 15740 15750 15760 15770 15780 15790 15800 15810 15820 15830 15840 15850 15860 15870 15880 15890 15900 15910 15920 15930 15940 15950 15960 15970 15980 15990 16000 16010 16020 16030 16040 16050 16060 16070 16080 16090 16100 16110 16120 16130 16140 16150 16160 16170 16180 16190 16200 16210 16220 16230 16240 16250 16260 16270 16280 16290 16300 16310 16320 16330 16340 16350 16360 16370 16380 16390 16400 16410 16420 16430 16440 16450 16460 16470 16480 16490 16500 16510 16520 16530 16540 16550 16560 16570 16580 16590 16600 16610 16620 16630 16640 16650 16660 16670 16680 16690 16700 16710 16720 16730 16740 16750 16760 16770 16780 16790 16800 16810 16820 16830 16840 16850 16860 16870 16880 16890 16900 16910 16920 16930 16940 16950 16960 16970 16980 16990 17000 17010 17020 17030 17040 17050 17060 17070 17080 17090 17100 17110 17120 17130 17140 17150 17160 17170 17180 17190 17200 17210 17220 17230 17240 17250 17260 17270 17280 17290 17300 17310 17320 17330 17340 17350 17360 17370 17380 17390 17400 17410 17420 17430 17440 17450 17460 17470 17480 17490 17500 17510 17520 17530 17540 17550 17560 17570 17580 17590 17600 17610 17620 17630 17640 17650 17660 17670 17680 17690 17700 17710 17720 17730 17740 17750 17760 17770 17780 17790 17800 17810 17820 17830 17840 17850 17860 17870 17880 17890 17900 17910 17920 17930 17940 17950 17960 17970 17980 17990 18000 18010 18020 18030 18040 18050 18060 18070 18080 18090 18100 18110 18120 18130 18140 18150 18160 18170 18180 18190 18200 18210 18220 18230 18240 18250 18260 18270 18280 18290 18300 18310 18320 18330 18340 18350 18360 18370 18380 18390 18400 18410 18420 18430 18440 18450 18460 18470 18480 18490 18500 18510 18520 18530 18540 18550 18560 18570 18580 18590 18600 18610 18620 18630 18640 18650 18660 18670 18680 18690 18700 18710 18720 18730 18740 18750 18760 18770 18780 18790 18800 18810 18820 18830 18840 18850 18860 18870 18880 18890 18900 18910 18920 18930 18940 18950 18960 18970 18980 18990 19000 19010 19020 19030 19040 19050 19060 19070 19080 19090 19100 19110 19120 19130 19140 19150 19160 19170 19180 19190 19200 19210 19220 19230 19240 19250 19260 19270 19280 19290 19300 19310 19320 19330 19340 19350 19360 19370 19380 19390 19400 19410 19420 19430 19440 19450 19460 19470 19480 19490 19500 19510 19520 19530 19540 19550 19560 19570 19580 19590 19600 19

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgyzstan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	KZ	Kazakhstan	SG	Singapore
CH	Switzerland	LJ	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovakia
CM	Cameroon	LR	Liberia	SN	Senegal
CN	China	LT	Lithuania	SZ	Swaziland
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	LV	Latvia	TG	Togo
DE	Germany	MC	Monaco	TJ	Tajikistan
DK	Denmark	MD	Republic of Moldova	TT	Trinidad and Tobago
EE	Estonia	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	UG	Uganda
FI	Finland	MN	Mongolia	US	United States of America
FR	France	MR	Mauritania	UZ	Uzbekistan
GA	Gabon			VN	Viet Nam

NUCLEOTIDE AND PROTEIN SEQUENCES OF  
VERTEBRATE SERRATE GENES AND METHODS BASED THEREON

This invention was made in part with government support under Grant numbers GM 29093 and NS 26084 awarded by the Department of Health and Human Services. The government has certain rights in the invention.

1. INTRODUCTION

10           The present invention relates to vertebrate Serrate genes and their encoded protein products, as well as derivatives and analogs thereof. Production of vertebrate Serrate proteins, derivatives, and antibodies is also provided. The invention further relates to therapeutic  
15 compositions and methods of diagnosis and therapy.

2. BACKGROUND OF THE INVENTION

Genetic analyses in *Drosophila* have been extremely useful in dissecting the complexity of developmental pathways  
20 and identifying interacting loci. However, understanding the precise nature of the processes that underlie genetic interactions requires a knowledge of the protein products of the genes in question.

Embryological, genetic and molecular evidence  
25 indicates that the early steps of ectodermal differentiation in *Drosophila* depend on cell interactions (Doe and Goodman, 1985, Dev. Biol. 111:206-219; Technau and Campos-Ortega, 1986, Dev. Biol. 195:445-454; Vässin et al., 1985, J. Neurogenet. 2:291-308; de la Concha et al., 1988, Genetics  
30 118:499-508; Xu et al., 1990, Genes Dev. 4:464-475; Artavanis-Tsakonas, 1988, Trends Genet. 4:95-100). Mutational analyses reveal a small group of zygotically-acting genes, the so called neurogenic loci, which affect the choice of ectodermal cells between epidermal and neural  
35 pathways (Poulson, 1937, Proc. Natl. Acad. Sci. 23:133-137; Lehmann et al., 1983, Wilhelm Roux's Arch. Dev. Biol. 192:62-74; Jürgens et al., 1984, Wilhelm Roux's Arch. Dev. Biol.



193:283-295; Wieschaus et al., 1984, Wilhelm Roux's Arch. Dev. Biol. 193:296-307; Nüsslein-Volhard et al., 1984, Wilhelm Roux's Arch. Dev. Biol. 193:267-282). Null mutations in any one of the zygotic neurogenic loci -- *Notch* (*N*), *Delta* (*Dl*), *mastermind* (*mam*), *Enhancer of Split* (*E(spl)*), *neuralized* (*neu*), and *big brain* (*bib*) -- result in hypertrophy of the nervous system at the expense of ventral and lateral epidermal structures. This effect is due to the misrouting of epidermal precursor cells into a neuronal pathway, and implies that neurogenic gene function is necessary to divert cells within the neurogenic region from a neuronal fate to an epithelial fate. *Serrate* has been identified as a genetic unit capable of interacting with the *Notch* locus (Xu et al., 1990, Genes Dev. 4:464-475). These genetic and developmental observations have led to the hypothesis that the protein products of the neurogenic loci function as components of a cellular interaction mechanism necessary for proper epidermal development (Artavanis-Tsakonas, S., 1988, Trends Genet. 4:95-100).

Mutational analyses also reveal that the action of the neurogenic genes is pleiotropic and is not limited solely to embryogenesis. For example, ommatidial, bristle and wing formation, which are known also to depend upon cell interactions, are affected by neurogenic mutations (Morgan et al., 1925, Bibliogr. Genet. 2:1-226; Welshons, 1956, Dros. Inf. Serv. 30:157-158; Preiss et al., 1988, EMBO J. 7:3917-3927; Shellenbarger and Mohler, 1978, Dev. Biol. 62:432-446; Technau and Campos-Ortega, 1986, Wilhelm Roux's Dev. Biol. 195:445-454; Tomlison and Ready, 1987, Dev. Biol. 120:366-376; Cagan and Ready, 1989, Genes Dev. 3:1099-1112).

Sequence analyses (Wharton et al., 1985, Cell 43:567-581; Kidd and Young, 1986, Mol. Cell. Biol. 6:3094-3108; Vässin, et al., 1987, EMBO J. 6:3431-3440; Kopczynski, et al., 1988, Genes Dev. 2:1723-1735) have shown that two of the neurogenic loci, *Notch* and *Delta*, appear to encode transmembrane proteins that span the membrane a single time. The *Notch* gene encodes a ~300 kd protein (we use "Notch" to

denote this protein) with a large N-terminal extracellular domain that includes 36 epidermal growth factor (EGF)-like tandem repeats followed by three other cysteine-rich repeats, designated *Notch/lin-12* repeats (Wharton, et al., 1985, Cell 5 43:567-581; Kidd and Young, 1986, Mol. Cell. Biol. 6:3094-3108; Yochem, et al., 1988, Nature 335:547-550). *Delta* encodes a ~100 kd protein (we use "Delta" to denote DL2M, the protein product of the predominant zygotic and maternal transcripts; Kopczynski, et al., 1988, Genes Dev. 2:1723-10 1735) that has nine EGF-like repeats within its extracellular domain (Vässin, et al., 1987, EMBO J. 6:3431-3440; Kopczynski, et al., 1988, Genes Dev. 2:1723-1735). Molecular studies have lead to the suggestion that Notch and Delta constitute biochemically interacting elements of a cell 15 communication mechanism involved in early developmental decisions (Fehon et al., 1990, Cell 61:523-534).

The EGF-like motif has been found in a variety of proteins, including those involved in the blood clotting cascade (Furie and Furie, 1988, Cell 53: 505-518). In 20 particular, this motif has been found in extracellular proteins such as the blood clotting factors IX and X (Rees et al., 1988, EMBO J. 7:2053-2061; Furie and Furie, 1988, Cell 53: 505-518), in other *Drosophila* genes (Knust et al., 1987 EMBO J. 761-766; Rothberg et al., 1988, Cell 55:1047-1059), 25 and in some cell-surface receptor proteins, such as thrombomodulin (Suzuki et al., 1987, EMBO J. 6:1891-1897) and LDL receptor (Sudhof et al., 1985, Science 228:815-822). A protein binding site has been mapped to the EGF repeat domain in thrombomodulin and urokinase (Kurosawa et al., 1988, J. 30 Biol. Chem 263:5993-5996; Appella et al., 1987, J. Biol. Chem. 262:4437-4440). The *Drosophila* *Serrate* gene has been cloned and characterized (PCT Publication WO 93/12141 dated June 24, 1993). However, prior to the present invention, despite attempts to achieve the same, no vertebrate *Serrate* 35 gene was available.

Citation of references hereinabove shall not be construed as an admission that such references are prior art to the present invention.

5

### 3. SUMMARY OF THE INVENTION

The present invention relates to nucleotide sequences of vertebrate Serrate genes (human Serrate and related genes of other species), and amino acid sequences of their encoded proteins, as well as derivatives (e.g.,  
10 fragments) and analogs thereof. Nucleic acids hybridizable to or complementary to the foregoing nucleotide sequences are also provided. In a specific embodiment, the Serrate protein is a human protein.

The invention relates to vertebrate Serrate  
15 derivatives and analogs of the invention which are functionally active, i.e., they are capable of displaying one or more known functional activities associated with a full-length (wild-type) Serrate protein. Such functional activities include but are not limited to antigenicity  
20 [ability to bind (or compete with Serrate for binding) to an anti-Serrate antibody], immunogenicity (ability to generate antibody which binds to Serrate), ability to bind (or compete with Serrate for binding) to Notch or other toporythmic proteins or fragments thereof ("adhesiveness"), ability to  
25 bind (or compete with Serrate for binding) to a receptor for Serrate. "Toporythmic proteins" as used herein, refers to the protein products of *Notch*, *Delta*, *Serrate*, *Enhancer of split*, and *Deltex*, as well as other members of this interacting gene family which may be identified, e.g., by  
30 virtue of the ability of their gene sequences to hybridize, or their homology to *Delta*, *Serrate*, or *Notch*, or the ability of their genes to display phenotypic interactions.

The invention further relates to fragments (and derivatives and analogs thereof) of vertebrate Serrate which  
35 comprise one or more domains of the Serrate protein, including but not limited to the intracellular domain, extracellular domain, transmembrane domain, membrane-

associated region, or one or more EGF-like (homologous) repeats of a Serrate protein, or any combination of the foregoing.

Antibodies to vertebrate Serrate, its derivatives 5 and analogs, are additionally provided.

Methods of production of the vertebrate Serrate proteins, derivatives and analogs, e.g., by recombinant means, are also provided.

The present invention also relates to therapeutic 10 and diagnostic methods and compositions based on vertebrate Serrate proteins and nucleic acids. The invention provides for treatment of disorders of cell fate or differentiation by administration of a therapeutic compound of the invention. Such therapeutic compounds (termed herein "Therapeutics") 15 include: vertebrate Serrate proteins and analogs and derivatives (including fragments) thereof; antibodies thereto; nucleic acids encoding the vertebrate Serrate proteins, analogs, or derivatives; and vertebrate Serrate antisense nucleic acids. In a preferred embodiment, a 20 Therapeutic of the invention is administered to treat a cancerous condition, or to prevent progression from a pre-neoplastic or non-malignant state into a neoplastic or a malignant state. In other specific embodiments, a Therapeutic of the invention is administered to treat a 25 nervous system disorder or to promote tissue regeneration and repair.

In one embodiment, Therapeutics which antagonize, or inhibit, Notch and/or Serrate function (hereinafter "Antagonist Therapeutics") are administered for therapeutic 30 effect. In another embodiment, Therapeutics which promote Notch and/or Serrate function (hereinafter "Agonist Therapeutics") are administered for therapeutic effect.

Disorders of cell fate, in particular hyperproliferative (e.g., cancer) or hypoproliferative 35 disorders, involving aberrant or undesirable levels of expression or activity or localization of Notch and/or

Serrate protein can be diagnosed by detecting such levels, as described more fully *infra*.

In a preferred aspect, a Therapeutic of the invention is a protein consisting of at least a fragment 5 (termed herein "adhesive fragment") of a vertebrate Serrate which mediates binding to a Notch protein or a fragment thereof.

### 3.1. DEFINITIONS

10 As used herein, underscoring or italicizing the name of a gene shall indicate the gene, in contrast to its encoded protein product which is indicated by the name of the gene in the absence of any underscoring. For example, "Serrate" shall mean the Serrate gene, whereas "Serrate" 15 shall indicate the protein product of the Serrate gene.

### 4. DESCRIPTION OF THE FIGURES

Figure 1. Nucleotide sequence (SEQ ID NO:1) and protein sequence (SEQ ID NO:2) of *Human Serrate-1* (also known 20 as *Human Jagged-1* (*HJ1*)).

Figure 2. "Complete" nucleotide sequence (SEQ ID NO:3) and amino acid sequence (SEQ ID NO:4) of *Human Serrate-2* (also known as *Human Jagged-2* (*HJ2*)) generated on the computer by combining the sequence of clones pBS15 and 25 pBS3-2 isolated from human fetal brain cDNA libraries. There is a deletion of approximately 120 nucleotides in the region of this sequence which encodes the portion of *Human Serrate-2* between the signal sequence and the beginning of the DSL domain.

30 Figure 3. Nucleotide sequence (SEQ ID NO:5) of chick Serrate (*C-Serrate*) cDNA.

Figure 4. Amino acid sequence (SEQ ID NO:6) of *C-Serrate* (lacking the amino-terminus of the signal sequence). The putative cleavage site following the signal 35 sequence (marking the predicted amino-terminus of the mature protein) is marked with an arrowhead; the DSL domain is indicated by asterisks; the EGF-like repeats (ELRs) are

underlined with dashed lines; the cysteine rich region between the ELRs and the transmembrane domain is marked between arrows, and the single transmembrane domain (between amino acids 1042 and 1066) is shown in bold.

5                   Figure 5. Alignment of the amino terminal sequences of *Drosophila melanogaster* Delta (SEQ ID NO:7) and Serrate (SEQ ID NO:8) with C-Serrate (SEQ ID NO:6). The region shown extends from the end of the signal sequence to the end of the DSL domain. The DSL domain is indicated.  
10 Identical amino acids in all three proteins are boxed.

                  Figure 6. Diagram showing the domain structures of *Drosophila* Delta and *Drosophila* Serrate compared with C-Serrate. The second cysteine-rich region just downstream of the EGF repeats, present only in C-Serrate and *Drosophila*  
15 Serrate, is not shown. Hydrophobic regions are shown in black; DSL domains are checkered and EGF-like repeats are hatched.

#### 5. DETAILED DESCRIPTION OF THE INVENTION

20                   The present invention relates to nucleotide sequences of vertebrate Serrate genes, and amino acid sequences of their encoded proteins. The invention further relates to fragments and other derivatives, and analogs, of vertebrate Serrate proteins. Nucleic acids encoding such  
25 fragments or derivatives are also within the scope of the invention. The invention provides vertebrate Serrate genes and their encoded proteins of many different species. The Serrate genes of the invention include human Serrate and related genes (homologs) in vertebrate species. In specific  
30 embodiments, the Serrate genes and proteins are from mammals. In a preferred embodiment of the invention, the Serrate protein is a human protein. In most preferred embodiments, the Serrate protein is Human Serrate-1 or Human Serrate-2. Production of the foregoing proteins and derivatives, e.g.,  
35 by recombinant methods, is provided.

                  The invention relates to vertebrate Serrate derivatives and analogs of the invention which ar

functionally active, i.e., they are capable of displaying one or more known functional activities associated with a full-length (wild-type) Serrate protein. Such functional activities include but are not limited to antigenicity  
5 [ability to bind (or compete with Serrate for binding) to an anti-Serrate antibody], immunogenicity (ability to generate antibody which binds to Serrate), ability to bind (or compete with Serrate for binding) to Notch or other toporythmic proteins or fragments thereof ("adhesiveness"), ability to  
10 bind (or compete with Serrate for binding) to a receptor for Serrate. "Toporythmic proteins" as used herein, refers to the protein products of *Notch*, *Delta*, *Serrate*, *Enhancer of split*, and *Deltex*, as well as other members of this interacting gene family which may be identified, e.g., by  
15 virtue of the ability of their gene sequences to hybridize, or their homology to *Delta*, *Serrate*, or *Notch*, or the ability of their genes to display phenotypic interactions.

The invention further relates to fragments (and derivatives and analogs thereof) of a vertebrate Serrate  
20 which comprise one or more domains of the Serrate protein, including but not limited to the intracellular domain, extracellular domain, transmembrane domain, membrane-associated region, or one or more EGF-like (homologous) repeats of a Serrate protein, or any combination of the  
25 foregoing.

Antibodies to Serrate, its derivatives and analogs, are additionally provided.

As demonstrated *infra*, Serrate plays a critical role in development and other physiological processes, in  
30 particular, as a ligand to Notch, which is involved in cell fate (differentiation) determination. In particular, Serrate is believed to play a major role in determining cell fates in the central nervous system. The nucleic acid and amino acid sequences and antibodies thereto of the invention can be used  
35 for the detection and quantitation of Serrate mRNA and protein of human and other species, to study expression thereof, to produce Serrate and fragments and other

derivatives and analogs thereof, in the study and manipulation of differentiation and other physiological processes. The present invention also relates to therapeutic and diagnostic methods and compositions based on Serrate proteins and nucleic acids. The invention provides for treatment of disorders of cell fate or differentiation by administration of a therapeutic compound of the invention. Such therapeutic compounds (termed herein "Therapeutics") include: vertebrate Serrate proteins and analogs and derivatives (including fragments) thereof; antibodies thereto; nucleic acids encoding the vertebrate Serrate proteins, analogs, or derivatives; and vertebrate Serrate antisense nucleic acids. In a preferred embodiment, a Therapeutic of the invention is administered to treat a cancerous condition, or to prevent progression from a pre-neoplastic or non-malignant state into a neoplastic or a malignant state. In other specific embodiments, a Therapeutic of the invention is administered to treat a nervous system disorder or to promote tissue regeneration and repair.

In one embodiment, Therapeutics which antagonize, or inhibit, Notch and/or Serrate function (hereinafter "Antagonist Therapeutics") are administered for therapeutic effect. In another embodiment, Therapeutics which promote Notch and/or Serrate function (hereinafter "Agonist Therapeutics") are administered for therapeutic effect.

Disorders of cell fate, in particular hyperproliferative (e.g., cancer) or hypoproliferative disorders, involving aberrant or undesirable levels of expression or activity or localization of Notch and/or Serrate protein can be diagnosed by detecting such levels, as described more fully *infra*.

In a preferred aspect, a Therapeutic of the invention is a protein consisting of at least a fragment (termed herein "adhesive fragment") of a vertebrate Serrate which mediates binding to a Notch protein or a fragment thereof.



The invention is illustrated by way of examples *infra* which disclose, *inter alia*, the cloning of a mouse Serrate homolog (Section 6), the cloning of a *Xenopus* (frog) Serrate homolog (Section 7), the cloning of a chick Serrate homolog (Section 8), and the cloning of the human Serrate homologs *Human Serrate-1* (HJ1) and *Human Serrate-2* (HJ2) (Section 9).

For clarity of disclosure, and not by way of limitation, the detailed description of the invention is divided into the sub-sections which follow.

### 5.1. ISOLATION OF THE SERRATE GENES

The invention relates to the nucleotide sequences of vertebrate Serrate nucleic acids. In specific embodiments, vertebrate Serrate nucleic acids comprise the cDNA sequences shown in Figure 1 (SEQ ID NO:1), Figure 2 (SEQ ID NO:3), Figure 3 (SEQ ID NO:6) or the coding regions thereof, or nucleic acids encoding a vertebrate Serrate protein (e.g., having the sequence of SEQ ID NO:2, 4, or 6).

The invention provides nucleic acids consisting of at least 8 nucleotides (i.e., a hybridizable portion) of a vertebrate Serrate sequence; in other embodiments, the nucleic acids consist of at least 10 (continuous) nucleotides, 25 nucleotides, 50 nucleotides, 100 nucleotides, 150 nucleotides, or 200 nucleotides of a vertebrate Serrate sequence, or a full-length vertebrate Serrate coding sequence. The invention also relates to nucleic acids hybridizable to or complementary to the foregoing sequences. In specific aspects, nucleic acids are provided which comprise a sequence complementary to at least 10, 25, 50, 100, or 200 nucleotides or the entire coding region of a Serrate gene.

In a specific embodiment, a nucleic acid which is hybridizable to a vertebrate Serrate nucleic acid (e.g., having sequence SEQ ID NO:1), or to a nucleic acid encoding a vertebrate Serrate derivative, under conditions of low stringency is provided. By way of example and not

limitation, procedures using such conditions of low stringency are as follows (see also Shilo and Weinberg, 1981, Proc. Natl. Acad. Sci. USA 78:6789-6792): Filters containing DNA are pretreated for 6 h at 40°C in a solution containing 5 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.1% PVP, 0.1% Ficoll, 1% BSA, and 500 µg/ml denatured salmon sperm DNA. Hybridizations are carried out in the same solution with the following modifications: 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 µg/ml salmon sperm DNA, 10% (wt/vol) 10 dextran sulfate, and 5-20 X 10<sup>6</sup> cpm <sup>32</sup>P-labeled probe is used. Filters are incubated in hybridization mixture for 18-20 h at 40°C, and then washed for 1.5 h at 55°C in a solution containing 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS. The wash solution is replaced with fresh solution 15 and incubated an additional 1.5 h at 60°C. Filters are blotted dry and exposed for autoradiography. If necessary, filters are washed for a third time at 65-68°C and reexposed to film. Other conditions of low stringency which may be used are well known in the art (e.g., as employed for cross- 20 species hybridizations).

In another specific embodiment, a nucleic acid which is hybridizable to a vertebrate Serrate nucleic acid under conditions of high stringency is provided. By way of example and not limitation, procedures using such conditions 25 of high stringency are as follows: Prehybridization of filters containing DNA is carried out for 8 h to overnight at 65°C in buffer composed of 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 µg/ml denatured salmon sperm DNA. Filters are hybridized for 48 h 30 at 65°C in prehybridization mixture containing 100 µg/ml denatured salmon sperm DNA and 5-20 X 10<sup>6</sup> cpm of <sup>32</sup>P-labeled probe. Washing of filters is done at 37°C for 1 h in a solution containing 2X SSC, 0.01% PVP, 0.01% Ficoll, and 0.01% BSA. This is followed by a wash in 0.1X SSC at 50°C 35 for 45 min before autoradiography. Other conditions of high stringency which may be used are well known in the art.

Nucleic acids encoding fragments and derivatives of vertebrate Serrate proteins (see Section 5.6), and vertebrate Serrate antisense nucleic acids (see Section 5.11) are additionally provided. As is readily apparent, as used  
5 herein, a "nucleic acid encoding a fragment or portion of a Serrate protein" shall be construed as referring to a nucleic acid encoding only the recited fragment or portion of the Serrate protein and not the other contiguous portions of the Serrate protein as a continuous sequence.

10 Fragments of vertebrate Serrate nucleic acids comprising regions of homology to other toporythmic proteins are also provided. The DSL regions (regions of homology with *Drosophila* Delta and Serrate) of Serrate proteins of other species are also provided. Nucleic acids encoding conserved  
15 regions between Delta and Serrate, such as those represented by Serrate amino acids 63-73, 124-134, 149-158, 195-206, 214-219, and 250-259 of SEQ ID NO:8, or by the DSL domains are also provided.

Specific embodiments for the cloning of a  
20 vertebrate Serrate gene, presented as a particular example but not by way of limitation, follows:

For expression cloning (a technique commonly known in the art), an expression library is constructed by methods known in the art. For example, mRNA (e.g., human) is  
25 isolated, cDNA is made and ligated into an expression vector (e.g., a bacteriophage derivative) such that it is capable of being expressed by the host cell into which it is then introduced. Various screening assays can then be used to select for the expressed Serrate product. In one embodiment,  
30 anti-Serrate antibodies can be used for selection.

In another preferred aspect, PCR is used to amplify the desired sequence in a genomic or cDNA library, prior to selection. Oligonucleotide primers representing known Serrate sequences can be used as primers in PCR. In a  
35 preferred aspect, the oligonucleotide primers encode at least part of the Serrate conserved segments of strong homology between Serrate and Delta. The synthetic oligonucleotides

may be utilized as primers to amplify by PCR sequences from a source (RNA or DNA), preferably a cDNA library, of potential interest. PCR can be carried out, e.g., by use of a Perkin-Elmer Cetus thermal cycler and Taq polymerase (Gene Amp<sup>®</sup>).

5 The DNA being amplified can include mRNA or cDNA or genomic DNA from any eukaryotic species. One can choose to synthesize several different degenerate primers, for use in the PCR reactions. It is also possible to vary the stringency of hybridization conditions used in priming the

10 PCR reactions, to allow for greater or lesser degrees of nucleotide sequence similarity between the known Serrate nucleotide sequence and the nucleic acid homolog being isolated. For cross species hybridization, low stringency conditions are preferred. For same species hybridization,

15 moderately stringent conditions are preferred. After successful amplification of a segment of a Serrate homolog, that segment may be cloned and sequenced, and utilized as a probe to isolate a complete cDNA or genomic clone. This, in turn, will permit the determination of the gene's complete

20 nucleotide sequence, the analysis of its expression, and the production of its protein product for functional analysis, as described *infra*. In this fashion, additional genes encoding Serrate proteins may be identified. Such a procedure is presented by way of example in various examples sections

25 *infra*.

The above-methods are not meant to limit the following general description of methods by which clones of vertebrate Serrate may be obtained.

Any vertebrate cell potentially can serve as the

30 nucleic acid source for the molecular cloning of the Serrate gene. The nucleic acid sequences encoding Serrate can be isolated from human, porcine, bovine, feline, avian, equine, canine, as well as additional primate sources, etc. For example, we have amplified fragments of the appropriate size

35 in mouse, *Xenopus*, and human, by PCR using cDNA libraries with *Drosophila* Serrate primers. The DNA may be obtained by standard procedures known in the art from cloned DNA (e.g., a

DNA "library"), by chemical synthesis, by cDNA cloning, or by the cloning of genomic DNA, or fragments thereof, purified from the desired cell. (See, for example, Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, 2d Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York; Glover, D.M. (ed.), 1985, DNA Cloning: A Practical Approach, MRL Press, Ltd., Oxford, U.K. Vol. I, II.) Clones derived from genomic DNA may contain regulatory and intron DNA regions in addition to coding regions; clones derived from cDNA will contain only exon sequences. Whatever the source, the gene should be molecularly cloned into a suitable vector for propagation of the gene.

In the molecular cloning of the gene from genomic DNA, DNA fragments are generated, some of which will encode the desired gene. The DNA may be cleaved at specific sites using various restriction enzymes. Alternatively, one may use DNase in the presence of manganese to fragment the DNA, or the DNA can be physically sheared, as for example, by sonication. The linear DNA fragments can then be separated according to size by standard techniques, including but not limited to, agarose and polyacrylamide gel electrophoresis and column chromatography.

Once the DNA fragments are generated, identification of the specific DNA fragment containing the desired gene may be accomplished in a number of ways. For example, if a Serrate (of any species) gene or its specific RNA, or a fragment thereof, e.g., an extracellular domain (see Section 5.6), is available and can be purified and labeled, the generated DNA fragments may be screened by nucleic acid hybridization to the labeled probe (Benton, W. and Davis, R., 1977, Science 196:180; Grunstein, M. And Hogness, D., 1975, Proc. Natl. Acad. Sci. U.S.A. 72:3961). Those DNA fragments with substantial homology to the probe will hybridize. It is also possible to identify the appropriate fragment by restriction enzyme digestion(s) and comparison of fragment sizes with those expected according to a known restriction map if such is available. Further

selection can be carried out on the basis of the properties of the gene. Alternatively, the presence of the gene may be detected by assays based on the physical, chemical, or immunological properties of its expressed product. For example, cDNA clones, or DNA clones which hybrid-select the proper mRNAs, can be selected which produce a protein that, e.g., has similar or identical electrophoretic migration, isoelectric focusing behavior, proteolytic digestion maps, receptor binding activity, *in vitro* aggregation activity ("adhesiveness") or antigenic properties as known for Serrate. If an antibody to Serrate is available, the Serrate protein may be identified by binding of labeled antibody to the putatively Serrate synthesizing clones, in an ELISA (enzyme-linked immunosorbent assay)-type procedure.

The Serrate gene can also be identified by mRNA selection by nucleic acid hybridization followed by *in vitro* translation. In this procedure, fragments are used to isolate complementary mRNAs by hybridization. Such DNA fragments may represent available, purified Serrate DNA of another species (e.g., human, chick). Immunoprecipitation analysis or functional assays (e.g., aggregation ability *in vitro*; binding to receptor; see *infra*) of the *in vitro* translation products of the isolated products of the isolated mRNAs identifies the mRNA and, therefore, the complementary DNA fragments that contain the desired sequences. In addition, specific mRNAs may be selected by adsorption of polysomes isolated from cells to immobilized antibodies specifically directed against Serrate protein. A radiolabeled Serrate cDNA can be synthesized using the selected mRNA (from the adsorbed polysomes) as a template. The radiolabeled mRNA or cDNA may then be used as a probe to identify the Serrate DNA fragments from among other genomic DNA fragments.

Alternatives to isolating the Serrate genomic DNA include, but are not limited to, chemically synthesizing the gene sequence itself from a known sequence or making cDNA to the mRNA which encodes the Serrate protein. For example, RNA

for cDNA cloning of the Serrate gene can be isolated from cells which express Serrate. Other methods are possible and within the scope of the invention.

The identified and isolated gene can then be  
5 inserted into an appropriate cloning vector. A large number of vector-host systems known in the art may be used. Possible vectors include, but are not limited to, plasmids or modified viruses, but the vector system must be compatible with the host cell used. Such vectors include, but are not  
10 limited to, bacteriophages such as lambda derivatives, or plasmids such as PBR322 or pUC plasmid derivatives. The insertion into a cloning vector can, for example, be accomplished by ligating the DNA fragment into a cloning vector which has complementary cohesive termini. However, if  
15 the complementary restriction sites used to fragment the DNA are not present in the cloning vector, the ends of the DNA molecules may be enzymatically modified. Alternatively, any site desired may be produced by ligating nucleotide sequences (linkers) onto the DNA termini; these ligated linkers may  
20 comprise specific chemically synthesized oligonucleotides encoding restriction endonuclease recognition sequences. In an alternative method, the cleaved vector and Serrate gene may be modified by homopolymeric tailing. Recombinant molecules can be introduced into host cells via  
25 transformation, transfection, infection, electroporation, etc., so that many copies of the gene sequence are generated.

In an alternative method, the desired gene may be identified and isolated after insertion into a suitable cloning vector in a "shot gun" approach. Enrichment for the  
30 desired gene, for example, by size fractionization, can be done before insertion into the cloning vector.

In specific embodiments, transformation of host cells with recombinant DNA molecules that incorporate the isolated Serrate gene, cDNA, or synthesized DNA sequence  
35 enables generation of multiple copies of the gene. Thus, the gene may be obtained in large quantities by growing transformants, isolating the recombinant DNA molecules from

the transformants and, when necessary, retrieving the inserted gene from the isolated recombinant DNA.

The Serrate sequences provided by the instant invention include those nucleotide sequences encoding  
5 substantially the same amino acid sequences as found in native Serrate proteins, and those encoded amino acid sequences with functionally equivalent amino acids, all as described in Section 5.6 *infra* for Serrate derivatives.

## 10 5.2. EXPRESSION OF THE SERRATE GENES

The nucleotide sequence coding for a vertebrate Serrate protein or a functionally active fragment or other derivative thereof (see Section 5.6), can be inserted into an appropriate expression vector, i.e., a vector which contains  
15 the necessary elements for the transcription and translation of the inserted protein-coding sequence. The necessary transcriptional and translational signals can also be supplied by the native vertebrate Serrate gene and/or its flanking regions. A variety of host-vector systems may be  
20 utilized to express the protein-coding sequence. These include but are not limited to mammalian cell systems infected with virus (e.g., vaccinia virus, adenovirus, etc.); insect cell systems infected with virus (e.g., baculovirus); microorganisms such as yeast containing yeast vectors, or  
25 bacteria transformed with bacteriophage, DNA, plasmid DNA, or cosmid DNA. The expression elements of vectors vary in their strengths and specificities. Depending on the host-vector system utilized, any one of a number of suitable transcription and translation elements may be used. In a  
30 specific embodiment, the adhesive portion of the Serrate gene is expressed. In other specific embodiments, a Human Serrate gene or a sequence encoding a functionally active portion of a human Serrate gene, such as Human Serrate-1 (HJ2) or Human Serrate-2 (HJ2), is expressed. In yet another embodiment, a  
35 fragment of Serrate comprising the extracellular domain, or other derivative, or analog of Serrate is expressed.



Any of the methods previously described for the insertion of DNA fragments into a vector may be used to construct expression vectors containing a chimeric gene consisting of appropriate transcriptional/translational control signals and the protein coding sequences. These methods may include *in vitro* recombinant DNA and synthetic techniques and *in vivo* recombinants (genetic recombination). Expression of nucleic acid sequence encoding a Serrate protein or peptide fragment may be regulated by a second nucleic acid sequence so that the Serrate protein or peptide is expressed in a host transformed with the recombinant DNA molecule. For example, expression of a Serrate protein may be controlled by any promoter/enhancer element known in the art. Promoters which may be used to control toporythmic gene expression include, but are not limited to, the SV40 early promoter region (Bernoist and Chambon, 1981, Nature 290:304-310), the promoter contained in the 3' long terminal repeat of Rous sarcoma virus (Yamamoto, et al., 1980, Cell 22:787-797), the herpes thymidine kinase promoter (Wagner et al., 1981, Proc. Natl. Acad. Sci. U.S.A. 78:1441-1445), the regulatory sequences of the metallothionein gene (Brinster et al., 1982, Nature 296:39-42); prokaryotic expression vectors such as the  $\beta$ -lactamase promoter (Villa-Kamaroff, et al., 1978, Proc. Natl. Acad. Sci. U.S.A. 75:3727-3731), or the *tac* promoter (DeBoer, et al., 1983, Proc. Natl. Acad. Sci. U.S.A. 80:21-25); see also "Useful proteins from recombinant bacteria" in Scientific American, 1980, 242:74-94; plant expression vectors comprising the nopaline synthetase promoter region (Herrera-Estrella et al., Nature 303:209-213) or the cauliflower mosaic virus 35S RNA promoter (Gardner, et al., 1981, Nucl. Acids Res. 9:2871), and the promoter of the photosynthetic enzyme ribulose biphosphate carboxylase (Herrera-Estrella et al., 1984, Nature 310:115-120); promoter elements from yeast or other fungi such as the Gal 4 promoter, the ADC (alcohol dehydrogenase) promoter, PGK (phosphoglycerol kinase) promoter, alkaline phosphatase promoter, and the following animal transcriptional control

regions, which exhibit tissue specificity and have been utilized in transgenic animals: elastase I gene control region which is active in pancreatic acinar cells (Swift et al., 1984, Cell 38:639-646; Ornitz et al., 1986, Cold Spring Harbor Symp. Quant. Biol. 50:399-409; MacDonald, 1987, Hepatology 7:425-515); insulin gene control region which is active in pancreatic beta cells (Hanahan, 1985, Nature 315:115-122), immunoglobulin gene control region which is active in lymphoid cells (Grosschedl et al., 1984, Cell 38:647-658; Adames et al., 1985, Nature 318:533-538; Alexander et al., 1987, Mol. Cell. Biol. 7:1436-1444), mouse mammary tumor virus control region which is active in testicular, breast, lymphoid and mast cells (Leder et al., 1986, Cell 45:485-495), albumin gene control region which is active in liver (Pinkert et al., 1987, Genes and Devel. 1:268-276), alpha-fetoprotein gene control region which is active in liver (Krumlauf et al., 1985, Mol. Cell. Biol. 5:1639-1648; Hammer et al., 1987, Science 235:53-58; alpha 1-antitrypsin gene control region which is active in the liver (Kelsey et al., 1987, Genes and Devel. 1:161-171), beta-globin gene control region which is active in myeloid cells (Mogram et al., 1985, Nature 315:338-340; Kollias et al., 1986, Cell 46:89-94; myelin basic protein gene control region which is active in oligodendrocyte cells in the brain (Readhead et al., 1987, Cell 48:703-712); myosin light chain-2 gene control region which is active in skeletal muscle (Sani, 1985, Nature 314:283-286), and gonadotropic releasing hormone gene control region which is active in the hypothalamus (Mason et al., 1986, Science 234:1372-1378).

Expression vectors containing Serrate gene inserts can be identified by three general approaches: (a) nucleic acid hybridization, (b) presence or absence of "marker" gene functions, and (c) expression of inserted sequences. In the first approach, the presence of a foreign gene inserted in an expression vector can be detected by nucleic acid hybridization using probes comprising sequences that are homologous to an inserted toporythmic gene. In the second

approach, the recombinant vector/host system can be identified and selected based upon the presence or absence of certain "marker" gene functions (e.g., thymidine kinase activity, resistance to antibiotics, transformation  
5 phenotype, occlusion body formation in baculovirus, etc.) caused by the insertion of foreign genes in the vector. For example, if the Serrate gene is inserted within the marker gene sequence of the vector, recombinants containing the Serrate insert can be identified by the absence of the marker  
10 gene function. In the third approach, recombinant expression vectors can be identified by assaying the foreign gene product expressed by the recombinant. Such assays can be based, for example, on the physical or functional properties of the Serrate gene product in vitro assay systems, e.g.,  
15 aggregation (binding) with Notch, binding to a receptor, binding with antibody.

Once a particular recombinant DNA molecule is identified and isolated, several methods known in the art may be used to propagate it. Once a suitable host system and  
20 growth conditions are established, recombinant expression vectors can be propagated and prepared in quantity. As previously explained, the expression vectors which can be used include, but are not limited to, the following vectors or their derivatives: human or animal viruses such as  
25 vaccinia virus or adenovirus; insect viruses such as baculovirus; yeast vectors; bacteriophage vectors (e.g., lambda), and plasmid and cosmid DNA vectors, to name but a few.

In addition, a host cell strain may be chosen which  
30 modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired. Expression from certain promoters can be elevated in the presence of certain inducers; thus, expression of the genetically engineered Serrate protein may  
35 be controlled. Furthermore, different host cells have characteristic and specific mechanisms for the translational and post-translational processing and modification (e.g.,

glycosylation, cleavage [e.g., of signal sequence]) of proteins. Appropriate cell lines or host systems can be chosen to ensure the desired modification and processing of the foreign protein expressed. For example, expression in a bacterial system can be used to produce an unglycosylated core protein product. Expression in yeast will produce a glycosylated product. Expression in mammalian cells can be used to ensure "native" glycosylation of a heterologous mammalian toporythmic protein. Furthermore, different vector/host expression systems may effect processing reactions such as proteolytic cleavages to different extents.

In other specific embodiments, the Serrate protein, fragment, analog, or derivative may be expressed as a fusion, or chimeric protein product (comprising the protein, fragment, analog, or derivative joined via a peptide bond to a heterologous protein sequence (of a different protein)). Such a chimeric product can be made by ligating the appropriate nucleic acid sequences encoding the desired amino acid sequences to each other by methods known in the art, in the proper coding frame, and expressing the chimeric product by methods commonly known in the art. Alternatively, such a chimeric product may be made by protein synthetic techniques, e.g., by use of a peptide synthesizer.

Both cDNA and genomic sequences can be cloned and expressed.

### 5.3. IDENTIFICATION AND PURIFICATION OF THE SERRATE GENE PRODUCTS

In particular aspects, the invention provides amino acid sequences of a vertebrate Serrate, preferably a human Serrate homolog, and fragments and derivatives thereof which comprise an antigenic determinant (i.e., can be recognized by an antibody) or which are otherwise functionally active, as well as nucleic acid sequences encoding the foregoing. "Functionally active" material as used herein refers to that material displaying one or more known functional activities associated with a full-length (wild-type) Serrate protein,

e.g., binding to Notch or a portion thereof, binding to any other Serrate ligand, antigenicity (binding to an anti-Serrate antibody), etc.

In specific embodiments, the invention provides  
5 fragments of a vertebrate Serrate protein consisting of at least 6 amino acids, 10 amino acids, 25 amino acids, 50 amino acids, or of at least 75 amino acids. In other embodiments, the proteins comprise or consist essentially of an  
10 extracellular domain, DSL domain, epidermal growth factor-like repeat (ELR) domain, one or any combination of ELRs, cysteine-rich region, transmembrane domain, or intracellular (cytoplasmic) domain, or a portion which binds to Notch, or any combination of the foregoing, of a Serrate protein. Fragments, or proteins comprising fragments, lacking some or  
15 all of the foregoing regions of a vertebrate Serrate protein are also provided. Nucleic acids encoding the foregoing are provided.

Once a recombinant which expresses the vertebrate Serrate gene sequence is identified, the gene product can be  
20 analyzed. This is achieved by assays based on the physical or functional properties of the product, including radioactive labelling of the product followed by analysis by gel electrophoresis, immunoassay, etc.

Once the Serrate protein is identified, it may be  
25 isolated and purified by standard methods including chromatography (e.g., ion exchange, affinity, and sizing column chromatography), centrifugation, differential solubility, or by any other standard technique for the purification of proteins. The functional properties may be  
30 evaluated using any suitable assay (see Section 5.7).

Alternatively, once a Serrate protein produced by a recombinant is identified, the amino acid sequence of the protein can be deduced from the nucleotide sequence of the chimeric gene contained in the recombinant. As a result, the  
35 protein can be synthesized by standard chemical methods known in the art (e.g., see Hunkapiller, M., et al., 1984, Nature 310:105-111).

In a specific embodiment of the present invention, such Serrate proteins, whether produced by recombinant DNA techniques or by chemical synthetic methods, include but are not limited to those containing, as a primary amino acid sequence, all or part of the amino acid sequence substantially as depicted in Figures 1, 2, or 3 (SEQ ID NO:2, 4, or 6, respectively), as well as fragments and other derivatives, and analogs thereof.

10           5.4.    STRUCTURE OF THE SERRATE GENES AND PROTEINS

The structure of the Serrate genes and proteins can be analyzed by various methods known in the art.

                  5.4.1.   GENETIC ANALYSIS

15           The cloned DNA or cDNA corresponding to the vertebrate Serrate gene can be analyzed by methods including but not limited to Southern hybridization (Southern, E.M., 1975, J. Mol. Biol. 98:503-517), Northern hybridization (see e.g., Freeman et al., 1983, Proc. Natl. Acad. Sci. U.S.A. 80:4094-4098), restriction endonuclease mapping (Maniatis, 20   T., 1982, Molecular Cloning, A Laboratory, Cold Spring Harbor, New York), and DNA sequence analysis. Polymerase chain reaction (PCR; U.S. Patent Nos. 4,683,202, 4,683,195 and 4,889,818; Gyllenstein et al., 1988, Proc. Natl. Acad. Sci. U.S.A. 85:7652-7656; Ochman et al., 1988, Genetics 25   120:621-623; Loh et al., 1989, Science 243:217-220) followed by Southern hybridization with a Serrate-specific probe can allow the detection of the Serrate gene in DNA from various cell types. Methods of amplification other than PCR are 30   commonly known and can also be employed. In one embodiment, Southern hybridization can be used to determine the genetic linkage of Serrate. Northern hybridization analysis can be used to determine the expression of the Serrate gene. Various cell types, at various states of development or 35   activity can be tested for Serrate expression. Examples of such techniques and their results are described in Section 6, *infra*. The stringency of the hybridization conditions for

both Southern and Northern hybridization can be manipulated to ensure detection of nucleic acids with the desired degree of relatedness to the specific Serrate probe used.

Restriction endonuclease mapping can be used to  
5 roughly determine the genetic structure of the Serrate gene. In a particular embodiment, cleavage with restriction enzymes can be used to derive the restriction map shown in Figure 2, *infra*. Restriction maps derived by restriction endonuclease cleavage can be confirmed by DNA sequence analysis.

10 DNA sequence analysis can be performed by any techniques known in the art, including but not limited to the method of Maxam and Gilbert (1980, Meth. Enzymol. 65:499-560), the Sanger dideoxy method (Sanger, F., et al., 1977, Proc. Natl. Acad. Sci. U.S.A. 74:5463), the use of T7 DNA  
15 polymerase (Tabor and Richardson, U.S. Patent No. 4,795,699), or use of an automated DNA sequenator (e.g., Applied Biosystems, Foster City, CA). The cDNA sequence of a representative Serrate gene comprises the sequence substantially as depicted in Figures 1 and 2, and is  
20 described in Section 9, *infra*.

#### 5.4.2. PROTEIN ANALYSIS

The amino acid sequence of the Serrate proteins can be derived by deduction from the DNA sequence, or  
25 alternatively, by direct sequencing of the protein, e.g., with an automated amino acid sequencer. The amino acid sequence of a representative Serrate protein comprises the sequence substantially as depicted in Figure 1, and detailed in Section 9, *infra*, with the representative mature protein  
30 that shown by amino acid numbers 30-1219.

The Serrate protein sequence can be further characterized by a hydrophilicity analysis (Hopp, T. and Woods, K., 1981, Proc. Natl. Acad. Sci. U.S.A. 78:3824). A hydrophilicity profile can be used to identify the  
35 hydrophobic and hydrophilic regions of the Serrate protein and the corresponding regions of the gene sequence which encode such regions.

Secondary, structural analysis (Chou, P. and Fasman, G., 1974, Biochemistry 13:222) can also be done, to identify regions of Serrate that assume specific secondary structures.

5 Manipulation, translation, and secondary structure prediction, as well as open reading frame prediction and plotting, can also be accomplished using computer software programs available in the art.

Other methods of structural analysis can also be  
10 employed. These include but are not limited to X-ray crystallography (Engstrom, A., 1974, Biochem. Exp. Biol. 11:7-13) and computer modeling (Fletterick, R. and Zoller, M. (eds.), 1986, Computer Graphics and Molecular Modeling, in Current Communications in Molecular Biology, Cold Spring  
15 Harbor Laboratory, Cold Spring Harbor, New York).

#### 5.5. GENERATION OF ANTIBODIES TO SERRATE PROTEINS AND DERIVATIVES THEREOF

According to the invention, a vertebrate Serrate  
20 protein, its fragments or other derivatives, or analogs thereof, may be used as an immunogen to generate antibodies which recognize such an immunogen. Such antibodies include but are not limited to polyclonal, monoclonal, chimeric, single chain, Fab fragments, and an Fab expression library.  
25 In a specific embodiment, antibodies to human Serrate are produced. In another embodiment, antibodies to the extracellular domain of Serrate are produced. In another embodiment, antibodies to the intracellular domain of Serrate are produced.

30 Various procedures known in the art may be used for the production of polyclonal antibodies to a Serrate protein or derivative or analog. In a particular embodiment, rabbit polyclonal antibodies to an epitope of the Serrate protein encoded by a sequence depicted in Figure 1, or a subsequence  
35 thereof, can be obtained. For the production of antibody, various host animals can be immunized by injection with the native Serrate protein, or a synthetic version, or derivative



(e.g., fragment) thereof, including but not limited to rabbits, mice, rats, etc. Various adjuvants may be used to increase the immunological response, depending on the host species, and including but not limited to Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanins, dinitrophenol, and potentially useful human adjuvants such as BCG (bacille Calmette-Guerin) and corynebacterium parvum.

For preparation of monoclonal antibodies directed toward a vertebrate Serrate protein sequence or analog thereof, any technique which provides for the production of antibody molecules by continuous cell lines in culture may be used. For example, the hybridoma technique originally developed by Kohler and Milstein (1975, Nature 256:495-497), as well as the trioma technique, the human B-cell hybridoma technique (Kozbor et al., 1983, Immunology Today 4:72), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole et al., 1985, in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96). In an additional embodiment of the invention, monoclonal antibodies can be produced in germ-free animals utilizing recent technology (PCT/US90/02545). According to the invention, human antibodies may be used and can be obtained by using human hybridomas (Cote et al., 1983, Proc. Natl. Acad. Sci. U.S.A. 80:2026-2030) or by transforming human B cells with EBV virus in vitro (Cole et al., 1985, in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, pp. 77-96). In fact, according to the invention, techniques developed for the production of "chimeric antibodies" (Morrison et al., 1984, Proc. Natl. Acad. Sci. U.S.A. 81:6851-6855; Neuberger et al., 1984, Nature 312:604-608; Takeda et al., 1985, Nature 314:452-454) by splicing the genes from a mouse antibody molecule specific for Serrate together with genes from a human antibody molecule of appropriate biological activity

can be used; such antibodies are within the scope of this invention.

According to the invention, techniques described for the production of single chain antibodies (U.S. Patent 5 4,946,778) can be adapted to produce Serrate-specific single chain antibodies. An additional embodiment of the invention utilizes the techniques described for the construction of Fab expression libraries (Huse et al., 1989, Science 246:1275-1281) to allow rapid and easy identification of monoclonal 10 Fab fragments with the desired specificity for Serrate proteins, derivatives, or analogs.

Antibody fragments which contain the idiotype of the molecule can be generated by known techniques. For example, such fragments include but are not limited to: the 15 F(ab')<sub>2</sub> fragment which can be produced by pepsin digestion of the antibody molecule; the Fab' fragments which can be generated by reducing the disulfide bridges of the F(ab')<sub>2</sub> fragment, and the Fab fragments which can be generated by treating the antibody molecule with papain and a reducing 20 agent.

In the production of antibodies, screening for the desired antibody can be accomplished by techniques known in the art, e.g. ELISA (enzyme-linked immunosorbent assay). For example, to select antibodies which recognize a specific 25 domain of a Serrate protein, one may assay generated hybridomas for a product which binds to a Serrate fragment containing such domain. For selection of an antibody specific to vertebrate (e.g., human) Serrate, one can select on the basis of positive binding to vertebrate Serrate and a 30 lack of binding to *Drosophila* Serrate. In another embodiment, one can select for binding to human Serrate and not to Serrate of other species.

The foregoing antibodies can be used in methods known in the art relating to the localization and activity of 35 the protein sequences of the invention (e.g., see Section 5.7, *infra*), e.g., for imaging these proteins, measuring

levels thereof in appropriate physiological samples, in diagnostic methods, etc.

Antibodies specific to a domain of a Serrate protein are also provided. In a specific embodiment, 5 antibodies which bind to a Notch-binding fragment of Serrate are provided.

In another embodiment of the invention (see *infra*), anti-Serrate antibodies and fragments thereof containing the binding domain are Therapeutics.

10

#### 5.6. SERRATE PROTEINS, DERIVATIVES AND ANALOGS

The invention further relates to vertebrate Serrate proteins, and derivatives (including but not limited to fragments) and analogs of Serrate proteins. Nucleic acids 15 encoding vertebrate Serrate protein derivatives and protein analogs are also provided. In one embodiment, the Serrate proteins are encoded by the vertebrate Serrate nucleic acids described in Section 5.1 *supra*. In particular aspects, the proteins, derivatives, or analogs are of frog, mouse, rat, 20 pig, cow, dog, monkey, or human Serrate proteins.

The production and use of derivatives and analogs related to vertebrate Serrate are within the scope of the present invention. In a specific embodiment, the derivative or analog is functionally active, i.e., capable of exhibiting 25 one or more functional activities associated with a full-length, wild-type Serrate protein. As one example, such derivatives or analogs which have the desired immunogenicity or antigenicity can be used, for example, in immunoassays, for immunization, for inhibition of Serrate activity, etc. 30 Such molecules which retain, or alternatively inhibit, a desired Serrate property, e.g., binding to Notch or other toporythmic proteins, binding to a cell-surface receptor, can be used as inducers, or inhibitors, respectively, of such property and its physiological correlates. A specific 35 embodiment relates to a Serrate fragment that can be bound by an anti-Serrate antibody but cannot bind to a Notch protein or other toporythmic protein. Derivatives or analogs of

Serrate can be tested for the desired activity by procedures known in the art, including but not limited to the assays described in Section 5.7.

In particular, Serrate derivatives can be made by  
5 altering Serrate sequences by substitutions, additions or deletions that provide for functionally equivalent molecules. Due to the degeneracy of nucleotide coding sequences, other DNA sequences which encode substantially the same amino acid sequence as a Serrate gene may be used in the practice of the  
10 present invention. These include but are not limited to nucleotide sequences comprising all or portions of Serrate genes which are altered by the substitution of different codons that encode a functionally equivalent amino acid residue within the sequence, thus producing a silent change.  
15 Likewise, the Serrate derivatives of the invention include, but are not limited to, those containing, as a primary amino acid sequence, all or part of the amino acid sequence of a Serrate protein including altered sequences in which functionally equivalent amino acid residues are substituted  
20 for residues within the sequence resulting in a silent change. For example, one or more amino acid residues within the sequence can be substituted by another amino acid of a similar polarity which acts as a functional equivalent, resulting in a silent alteration. Substitutes for an amino  
25 acid within the sequence may be selected from other members of the class to which the amino acid belongs. For example, the nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan and methionine. The polar neutral amino acids  
30 include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine. The positively charged (basic) amino acids include arginine, lysine and histidine. The negatively charged (acidic) amino acids include aspartic acid and glutamic acid.

35 In a specific embodiment of the invention, proteins consisting of or comprising a fragment of a vertebrate Serrate protein consisting of at least 10 (continuous) amino

acids of the Serrate protein is provided. In other embodiments, the fragment consists of at least 20 or 50 amino acids of the Serrate protein. In specific embodiments, such fragments are not larger than 35, 100 or 200 amino acids.

5 Derivatives or analogs of vertebrate Serrate include but are not limited to those peptides which are substantially homologous to a vertebrate Serrate or a fragment thereof (e.g., at least 30% identity over an amino acid sequence of identical size) or whose encoding nucleic acid is capable of

10 hybridizing to a coding vertebrate Serrate sequence.

The Serrate derivatives and analogs of the invention can be produced by various methods known in the art. The manipulations which result in their production can occur at the gene or protein level. For example, the cloned

15 Serrate gene sequence can be modified by any of numerous strategies known in the art (Maniatis, T., 1990, Molecular Cloning, A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York). The sequence can be cleaved at appropriate sites with restriction

20 endonuclease(s), followed by further enzymatic modification if desired, isolated, and ligated *in vitro*. In the production of the gene encoding a derivative or analog of Serrate, care should be taken to ensure that the modified gene remains within the same translational reading frame as

25 Serrate, uninterrupted by translational stop signals, in the gene region where the desired Serrate activity is encoded.

Additionally, the Serrate-encoding nucleic acid sequence can be mutated *in vitro* or *in vivo*, to create and/or destroy translation, initiation, and/or termination

30 sequences, or to create variations in coding regions and/or form new restriction endonuclease sites or destroy preexisting ones, to facilitate further *in vitro* modification. Any technique for mutagenesis known in the art can be used, including but not limited to, *in vitro* site-

35 directed mutagenesis (Hutchinson, C., et al., 1978, J. Biol. Chem 253:6551), use of TAB linkers (Pharmacia), etc.

Manipulations of the Serrate sequence may also be made at the protein level. Included within the scope of the invention are Serrate protein fragments or other derivatives or analogs which are differentially modified during or after translation, e.g., by glycosylation, acetylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, linkage to an antibody molecule or other cellular ligand, etc. Any of numerous chemical modifications may be carried out by known techniques, including but not limited to specific chemical cleavage by cyanogen bromide, trypsin, chymotrypsin, papain, V8 protease,  $\text{NaBH}_4$ ; acetylation, formylation, oxidation, reduction; metabolic synthesis in the presence of tunicamycin; etc.

In addition, analogs and derivatives of Serrate can be chemically synthesized. For example, a peptide corresponding to a portion of a Serrate protein which comprises the desired domain (see Section 5.6.1), or which mediates the desired aggregation activity *in vitro*, or binding to a receptor, can be synthesized by use of a peptide synthesizer. Furthermore, if desired, nonclassical amino acids or chemical amino acid analogs can be introduced as a substitution or addition into the Serrate sequence. Non-classical amino acids include but are not limited to the D-isomers of the common amino acids,  $\alpha$ -amino isobutyric acid, 4-aminobutyric acid, hydroxyproline, sarcosine, citrulline, cysteic acid, t-butylglycine, t-butylalanine, phenylglycine, cyclohexylalanine,  $\beta$ -alanine, designer amino acids such as  $\beta$ -methyl amino acids,  $\text{Ca}$ -methyl amino acids, and  $\text{Na}$ -methyl amino acids.

In a specific embodiment, the Serrate derivative is a chimeric, or fusion, protein comprising a vertebrate Serrate protein or fragment thereof (preferably consisting of at least a domain or motif of the Serrate protein, or at least 10 amino acids of the Serrate protein) joined at its amino- or carboxy-terminus via a peptide bond to an amino acid sequence of a different protein. In one embodiment,

such a chimeric protein is produced by recombinant expression of a nucleic acid encoding the protein (comprising a Serrate-coding sequence joined in-frame to a coding sequence for a different protein). Such a chimeric product can be made by  
5 ligating the appropriate nucleic acid sequences encoding the desired amino acid sequences to each other by methods known in the art, in the proper coding frame, and expressing the chimeric product by methods commonly known in the art. Alternatively, such a chimeric product may be made by protein  
10 synthetic techniques, e.g., by use of a peptide synthesizer. In a specific embodiment, a chimeric nucleic acid encoding a mature vertebrate Serrate protein with a heterologous signal sequence is expressed such that the chimeric protein is expressed and processed by the cell to the mature Serrate  
15 protein. As another example, and not by way of limitation, a recombinant molecule can be constructed according to the invention, comprising coding portions of both Serrate and another toporythmic gene, e.g., Delta. The encoded protein of such a recombinant molecule could exhibit properties  
20 associated with both Serrate and Delta and portray a novel profile of biological activities, including agonists as well as antagonists. The primary sequence of Serrate and Delta may also be used to predict tertiary structure of the molecules using computer simulation (Hopp and Woods, 1981,  
25 Proc. Natl. Acad. Sci. U.S.A. 78:3824-3828); Serrate/Delta chimeric recombinant genes could be designed in light of correlations between tertiary structure and biological function. Likewise, chimeric genes comprising portions of a vertebrate Serrate fused to any heterologous protein-encoding  
30 sequences may be constructed. A specific embodiment relates to a chimeric protein comprising a fragment of a vertebrate Serrate of at least ten amino acids.

In another specific embodiment, the Serrate derivative is a fragment of Serrate comprising a region of  
35 homology with another toporythmic protein. As used herein, a region of a first protein shall be considered "homologous" to a second protein when the amino acid sequence of the region

is at least 30% identical or at least 75% either identical or involving conservative changes, when compared to any sequence in the second protein of an equal number of amino acids as the number contained in the region. For example, such a  
5 Serrate fragment can comprise one or more regions homologous to Delta, or DSL domains or portions thereof.

Other specific embodiments of derivatives and analogs are described in the subsections below and examples sections *infra*.

10

#### 5.6.1. DERIVATIVES OF SERRATE CONTAINING ONE OR MORE DOMAINS OF THE PROTEIN

In a specific embodiment, the invention relates to vertebrate Serrate derivatives and analogs, in particular  
15 vertebrate Serrate fragments and derivatives of such fragments, that comprise, or alternatively consist of, one or more domains of the Serrate protein, including but not limited to the extracellular domain, DSL domain, ELR domain, cysteine rich domain, transmembrane domain, intracellular  
20 domain, membrane-associated region, and one or more of the EGF-like repeats (ELR) of the Serrate protein, or any combination of the foregoing. In particular examples relating to the human and chick Serrate proteins, such domains are identified in Examples Section 9 and 8,  
25 respectively.

In a specific embodiment, the molecules comprising specific fragments of vertebrate Serrate are those comprising fragments in the respective Serrate protein most homologous to specific fragments of the *Drosophila* Serrate and/or Delta  
30 proteins. In particular embodiments, such a molecule comprises or consists of the amino acid sequences homologous to SEQ ID NO:10, 12, or 18. Alternatively, a fragment comprising a domain of a Serrate homolog can be identified by protein analysis methods as described in Section 5.3.2.

35



#### 5.6.2. DERIVATIVES OF SERRATE THAT MEDIATE BINDING TO TOPORYTHMIC PROTEIN DOMAINS

The invention also provides for vertebrate Serrate fragments, and analogs or derivatives of such fragments, which mediate binding to toporythmic proteins (and thus are termed herein "adhesive"), and nucleic acid sequences encoding the foregoing.

In a specific embodiment, the adhesive fragment of Serrate is that comprising the portion of Serrate most homologous to about amino acid numbers 85-283 or 79-282 of the *Drosophila* Serrate sequence (see PCT Publication WO 93/12141 dated June 24, 1993).

In a particular embodiment, the adhesive fragment of a Serrate protein comprises the DSL domain, or a portion thereof. Subfragments within the DSL domain that mediate binding to Notch can be identified by analysis of constructs expressing deletion mutants.

The ability to bind to a toporythmic protein (preferably Notch) can be demonstrated by *in vitro* aggregation assays with cells expressing such a toporythmic protein as well as cells expressing Serrate or a Serrate derivative (See Section 5.7). That is, the ability of a Serrate fragment to bind to a Notch protein can be demonstrated by detecting the ability of the Serrate fragment, when expressed on the surface of a first cell, to bind to a Notch protein expressed on the surface of a second cell.

The nucleic acid sequences encoding toporythmic proteins or adhesive domains thereof, for use in such assays, can be isolated from human, porcine, bovine, feline, avian, equine, canine, or insect, as well as primate sources and any other species in which homologs of known toporythmic genes can be identified.

### 5.7. ASSAYS OF SERRATE PROTEINS, DERIVATIVES AND ANALOGS

The functional activity of vertebrate Serrate proteins, derivatives and analogs can be assayed by various  
5 methods.

For example, in one embodiment, where one is assaying for the ability to bind or compete with wild-type Serrate for binding to anti-Serrate antibody, various immunoassays known in the art can be used, including but not  
10 limited to competitive and non-competitive assay systems using techniques such as radioimmunoassays, ELISA (enzyme linked immunosorbent assay), "sandwich" immunoassays, immunoradiometric assays, gel diffusion precipitin reactions, immunodiffusion assays, *in situ* immunoassays (using colloidal  
15 gold, enzyme or radioisotope labels, for example), western blots, precipitation reactions, agglutination assays (e.g., gel agglutination assays, hemagglutination assays), complement fixation assays, immunofluorescence assays, protein A assays, and immunoelectrophoresis assays, etc. In  
20 one embodiment, antibody binding is detected by detecting a label on the primary antibody. In another embodiment, the primary antibody is detected by detecting binding of a secondary antibody or reagent to the primary antibody. In a further embodiment, the secondary antibody is labeled. Many  
25 means are known in the art for detecting binding in an immunoassay and are within the scope of the present invention.

In another embodiment, where one is assaying for the ability to mediate binding to a toporythmic protein,  
30 e.g., Notch, one can carry out an *in vitro* aggregation assay such as described in PCT Publication WO 93/12141 dated June 24, 1993 (see also Fehon et al., 1990, Cell 61:523-534; Rebay et al., 1991, Cell 67:687-699).

In another embodiment, where a receptor for Serrate  
35 is identified, receptor binding can be assayed, e.g., by means well-known in the art. In another embodiment, physiological correlates of Serrate binding to cells

expressing a Serrate receptor (signal transduction) can be assayed.

In another embodiment, in insect or other model systems, genetic studies can be done to study the phenotypic effect of a Serrate mutant that is a derivative or analog of wild-type vertebrate Serrate.

Other methods will be known to the skilled artisan and are within the scope of the invention.

10

#### 5.8. THERAPEUTIC USES

The invention provides for treatment of disorders of cell fate or differentiation by administration of a therapeutic compound of the invention. Such therapeutic compounds (termed herein "Therapeutics") include: vertebrate Serrate proteins and analogs and derivatives (including fragments) thereof (e.g., as described hereinabove); antibodies thereto (as described hereinabove); nucleic acids encoding the vertebrate Serrate proteins, analogs, or derivatives (e.g., as described hereinabove); and Serrate antisense nucleic acids. As stated *supra*, the Antagonist Therapeutics of the invention are those Therapeutics which antagonize, or inhibit, a vertebrate Serrate function and/or Notch function (since Serrate is a Notch ligand). Such Antagonist Therapeutics are most preferably identified by use of known convenient *in vitro* assays, e.g., based on their ability to inhibit binding of Serrate to another protein (e.g., a Notch protein), or inhibit any known Notch or Serrate function as preferably assayed *in vitro* or in cell culture, although genetic assays (e.g., in *Drosophila*) may also be employed. In a preferred embodiment, the Antagonist Therapeutic is a protein or derivative thereof comprising a functionally active fragment such as a fragment of Serrate which mediates binding to Notch, or an antibody thereto. In other specific embodiments, such an Antagonist Therapeutic is a nucleic acid capable of expressing a molecule comprising a fragment of Serrate which binds to Notch, or a Serrate antisense nucleic acid (see Section 5.11 herein). It should

be noted that preferably, suitable *in vitro* or *in vivo* assays, as described *infra*, should be utilized to determine the effect of a specific Therapeutic and whether its administration is indicated for treatment of the affected  
5 tissue, since the developmental history of the tissue may determine whether an Antagonist or Agonist Therapeutic is desired.

In addition, the mode of administration, e.g., whether administered in soluble form or administered via its  
10 encoding nucleic acid for intracellular recombinant expression, of the Serrate protein or derivative can affect whether it acts as an agonist or antagonist.

In another embodiment of the invention, a nucleic acid containing a portion of a vertebrate Serrate gene is  
15 used, as an Antagonist Therapeutic, to promote Serrate inactivation by homologous recombination (Koller and Smithies, 1989, Proc. Natl. Acad. Sci. USA 86:8932-8935; Zijlstra et al., 1989, Nature 342:435-438).

The Agonist Therapeutics of the invention, as  
20 described *supra*, promote Serrate function. Such Agonist Therapeutics include but are not limited to proteins and derivatives comprising the portions of Notch that mediate binding to Serrate, and nucleic acids encoding the foregoing (which can be administered to express their encoded products  
25 *in vivo*).

Further descriptions and sources of Therapeutics of the inventions are found in Sections 5.1 through 5.7 herein.

Molecules which retain, or alternatively inhibit, a desired Serrate property, e.g., binding to Notch, binding to  
30 an intracellular ligand, can be used therapeutically as inducers, or inhibitors, respectively, of such property and its physiological correlates. In a specific embodiment, a peptide (e.g., in the range of 10-50 or 15-25 amino acids; and particularly of about 10, 15, 20 or 25 amino acids)  
35 containing the sequence of a portion of a vertebrate Serrate which binds to Notch is used to antagonize Notch function. In a specific embodiment, such an Antagonist Therapeutic is

used to treat or prevent human or other malignancies associated with increased Notch expression (e.g., cervical cancer, colon cancer, breast cancer, squamous adenocarcinomas (see *infra*)). Derivatives or analogs of Serrate can be tested for the desired activity by procedures known in the art, including but not limited to the assays described in the examples *infra*. For example, molecules comprising vertebrate Serrate fragments which bind to Notch EGF-repeats (ELR) 11 and 12 and which are smaller than a DSL domain, can be obtained and selected by expressing deletion mutants and assaying for binding of the expressed product to Notch by any of the several methods (e.g., *in vitro* cell aggregation assays, interaction trap system), some of which are described in the Examples Sections *infra*. In one specific embodiment, peptide libraries can be screened to select a peptide with the desired activity; such screening can be carried out by assaying, e.g., for binding to Notch or a molecule containing the Notch ELR 11 and 12 repeats.

The Agonist and Antagonist Therapeutics of the invention have therapeutic utility for disorders of cell fate. The Agonist Therapeutics are administered therapeutically (including prophylactically): (1) in diseases or disorders involving an absence or decreased (relative to normal, or desired) levels of Notch or Serrate function, for example, in patients where Notch or Serrate protein is lacking, genetically defective, biologically inactive or underactive, or underexpressed; and (2) in diseases or disorders wherein *in vitro* (or *in vivo*) assays (see *infra*) indicate the utility of Serrate agonist administration. The absence or decreased levels in Notch or Serrate function can be readily detected, e.g., by obtaining a patient tissue sample (e.g., from biopsy tissue) and assaying it *in vitro* for protein levels, structure and/or activity of the expressed Notch or Serrate protein. Many methods standard in the art can be thus employed, including but not limited to immunoassays to detect and/or visualize Notch or Serrate protein (e.g., Western blot, immunoprecipitation followed by

sodium dodecyl sulfate polyacrylamide gel electrophoresis, immunocytochemistry, etc.) and/or hybridization assays to detect Notch or Serrate expression by detecting and/or visualizing respectively Notch or Serrate mRNA (e.g., Northern assays, dot blots, *in situ* hybridization, etc.)

*In vitro* assays which can be used to determine whether administration of a specific Agonist Therapeutic or Antagonist Therapeutic is indicated, include *in vitro* cell culture assays in which a patient tissue sample is grown in culture, and exposed to or otherwise administered a Therapeutic, and the effect of such Therapeutic upon the tissue sample is observed. In one embodiment, where the patient has a malignancy, a sample of cells from such malignancy is plated out or grown in culture, and the cells are then exposed to a Therapeutic. A Therapeutic which inhibits survival or growth of the malignant cells (e.g., by promoting terminal differentiation) is selected for therapeutic use *in vivo*. Many assays standard in the art can be used to assess such survival and/or growth; for example, cell proliferation can be assayed by measuring <sup>3</sup>H-thymidine incorporation, by direct cell count, by detecting changes in transcriptional activity of known genes such as proto-oncogenes (e.g., *fos*, *myc*) or cell cycle markers; cell viability can be assessed by trypan blue staining, differentiation can be assessed visually based on changes in morphology, etc. In a specific aspect, the malignant cell cultures are separately exposed to (1) an Agonist Therapeutic, and (2) an Antagonist Therapeutic; the result of the assay can indicate which type of Therapeutic has therapeutic efficacy.

In another embodiment, a Therapeutic is indicated for use which exhibits the desired effect, inhibition or promotion of cell growth, upon a patient cell sample from tissue having or suspected of having a hyper- or hypoproliferative disorder, respectively. Such hyper- or hypoproliferative disorders include but are not limited to those described in Sections 5.8.1 through 5.8.3 *infra*.

In another specific embodiment, a Therapeutic is indicated for use in treating nerve injury or a nervous system degenerative disorder (see Section 5.8.2) which exhibits *in vitro* promotion of nerve regeneration/neurite extension from nerve cells of the affected patient type.

In addition, administration of an Antagonist Therapeutic of the invention is also indicated in diseases or disorders determined or known to involve a Notch or Serrate dominant activated phenotype ("gain of function" mutations.)

10 Administration of an Agonist Therapeutic is indicated in diseases or disorders determined or known to involve a Notch or Serrate dominant negative phenotype ("loss of function" mutations). The functions of various structural domains of the Notch protein have been investigated *in vivo*, by

15 ectopically expressing a series of *Drosophila* Notch deletion mutants under the hsp70 heat-shock promoter, as well as eye-specific promoters (see Rebay et al., 1993, Cell 74:319-329). Two classes of dominant phenotypes were observed, one

20 of Notch gain-of-function mutations. Dominant "activated" phenotypes resulted from overexpression of a protein lacking most extracellular sequences, while dominant "negative" phenotypes resulted from overexpression of a protein lacking most intracellular sequences. The results indicated that

25 Notch functions as a receptor whose extracellular domain mediates ligand-binding, resulting in the transmission of developmental signals by the cytoplasmic domain. We have shown that Serrate binds to the Notch ELR 11 and 12 (see PCT Publication WO 93/12141).

30 In various specific embodiments, *in vitro* assays can be carried out with representative cells of cell types involved in a patient's disorder, to determine if a Therapeutic has a desired effect upon such cell types.

In another embodiment, cells of a patient tissue sample suspected of being pre-neoplastic are similarly plated out or grown *in vitro*, and exposed to a Therapeutic. The Therapeutic which results in a cell phenotype that is more

normal (i.e., less representative of a pre-neoplastic state, neoplastic state, malignant state, or transformed phenotype) is selected for therapeutic use. Many assays standard in the art can be used to assess whether a pre-neoplastic state, 5 neoplastic state, or a transformed or malignant phenotype, is present. For example, characteristics associated with a transformed phenotype (a set of *in vitro* characteristics associated with a tumorigenic ability *in vivo*) include a more rounded cell morphology, looser substratum attachment, loss 10 of contact inhibition, loss of anchorage dependence, release of proteases such as plasminogen activator, increased sugar transport, decreased serum requirement, expression of fetal antigens, disappearance of the 250,000 dalton surface protein, etc. (see Luria et al., 1978, *General Virology*, 3d 15 Ed., John Wiley & Sons, New York pp. 436-446).

In other specific embodiments, the *in vitro* assays described supra can be carried out using a cell line, rather than a cell sample derived from the specific patient to be treated, in which the cell line is derived from or displays 20 characteristic(s) associated with the malignant, neoplastic or pre-neoplastic disorder desired to be treated or prevented, or is derived from the neural or other cell type upon which an effect is desired, according to the present invention.

25 The Antagonist Therapeutics are administered therapeutically (including prophylactically): (1) in diseases or disorders involving increased (relative to normal, or desired) levels of Notch or Serrate function, for example, where the Notch or Serrate protein is overexpressed or 30 overactive; and (2) in diseases or disorders wherein *in vitro* (or *in vivo*) assays indicate the utility of Serrate antagonist administration. The increased levels of Notch or Serrate function can be readily detected by methods such as those described above, by quantifying protein and/or RNA. In 35 *in vitro* assays with cells of patient tissue sample or the appropriate cell line or cell type, to determine therapeutic utility, can be carried out as described above.



5.8.1. MALIGNANCIES

Malignant and pre-neoplastic conditions which can be tested as described supra for efficacy of intervention with Antagonist or Agonist Therapeutics, and which can be  
 5 treated upon thus observing an indication of therapeutic utility, include but are not limited to those described below in Sections 5.8.1 and 5.9.1.

Malignancies and related disorders, cells of which type can be tested *in vitro* (and/or *in vivo*), and upon  
 10 observing the appropriate assay result, treated according to the present invention, include but are not limited to those listed in Table 1 (for a review of such disorders, see Fishman et al., 1985, *Medicine*, 2d Ed., J.B. Lippincott Co., Philadelphia):

15

---

TABLE 1  
MALIGNANCIES AND RELATED DISORDERS

20	Leukemia
	acute leukemia
25	acute lymphocytic leukemia
	acute myelocytic leukemia
	myeloblastic
	promyelocytic
	myelomonocytic
	monocytic
	erythroleukemia
30	chronic leukemia
	chronic myelocytic (granulocytic) leukemia
	chronic lymphocytic leukemia
	Polycythemia vera
	Lymphoma
	Hodgkin's disease
	non-Hodgkin's disease
35	Multiple myeloma
	Waldenström's macroglobulinemia
	Heavy chain disease
	Solid tumors
	sarcomas and carcinomas
	fibrosarcoma
	myxosarcoma
	liposarcoma
	chondrosarcoma
	osteogenic sarcoma
	chordoma

angiosarcoma  
endotheliosarcoma  
lymphangiosarcoma  
lymphangioendotheliosarcoma  
synovioma  
5 mesothelioma  
Ewing's tumor  
leiomyosarcoma  
rhabdomyosarcoma  
colon carcinoma  
pancreatic cancer  
breast cancer  
ovarian cancer  
10 prostate cancer  
squamous cell carcinoma  
basal cell carcinoma  
adenocarcinoma  
sweat gland carcinoma  
sebaceous gland carcinoma  
papillary carcinoma  
15 papillary adenocarcinomas  
cystadenocarcinoma  
medullary carcinoma  
bronchogenic carcinoma  
renal cell carcinoma  
hepatoma  
bile duct carcinoma  
choriocarcinoma  
20 seminoma  
embryonal carcinoma  
Wilms' tumor  
cervical cancer  
testicular tumor  
lung carcinoma  
small cell lung carcinoma  
25 bladder carcinoma  
epithelial carcinoma  
glioma  
astrocytoma  
medulloblastoma  
craniopharyngioma  
ependymoma  
pinealoma  
30 hemangioblastoma  
acoustic neuroma  
oligodendroglioma  
menangioma  
melanoma  
neuroblastoma  
retinoblastoma

35

In specific embodiments, malignancy or dysproliferative changes (such as metaplasias and dysplasias) are treated or prevented in epithelial tissues such as those in the cervix, esophagus, and lung.

5 Malignancies of the colon and cervix exhibit increased expression of human Notch relative to such non-malignant tissue (see PCT Publication no. WO 94/07474 published April 14, 1994, incorporated by reference herein in its entirety). Thus, in specific embodiments, malignancies  
10 or premalignant changes of the colon or cervix are treated or prevented by administering an effective amount of an Antagonist Therapeutic, e.g., a Serrate derivative, that antagonizes Notch function. The presence of increased Notch expression in colon, and cervical cancer suggests that many  
15 more cancerous and hyperproliferative conditions exhibit upregulated Notch. Thus, in specific embodiments, various cancers, e.g., breast cancer, squamous adenocarcinoma, seminoma, melanoma, and lung cancer, and premalignant changes therein, as well as other hyperproliferative disorders, can  
20 be treated or prevented by administration of an Antagonist Therapeutic that antagonizes Notch function.

#### 5.8.2. NERVOUS SYSTEM DISORDERS

Nervous system disorders, involving cell types  
25 which can be tested as described *supra* for efficacy of intervention with Antagonist or Agonist Therapeutics, and which can be treated upon thus observing an indication of therapeutic utility, include but are not limited to nervous system injuries, and diseases or disorders which result in  
30 either a disconnection of axons, a diminution or degeneration of neurons, or demyelination. Nervous system lesions which may be treated in a patient (including human and non-human mammalian patients) according to the invention include but are not limited to the following lesions of either the  
35 central (including spinal cord, brain) or peripheral nervous systems:

- 5 (i) traumatic lesions, including lesions caused by physical injury or associated with surgery, for example, lesions which sever a portion of the nervous system, or compression injuries;
- (ii) ischemic lesions, in which a lack of oxygen in a portion of the nervous system results in neuronal injury or death, including cerebral infarction or ischemia, or spinal cord infarction or ischemia;
- 10 (iii) malignant lesions, in which a portion of the nervous system is destroyed or injured by malignant tissue which is either a nervous system associated malignancy or a malignancy derived from non-nervous system tissue;
- 15 (iv) infectious lesions, in which a portion of the nervous system is destroyed or injured as a result of infection, for example, by an abscess or associated with infection by human immunodeficiency virus, herpes zoster, or
- 20 herpes simplex virus or with Lyme disease, tuberculosis, syphilis;
- (v) degenerative lesions, in which a portion of the nervous system is destroyed or injured as a result of a degenerative process including
- 25 but not limited to degeneration associated with Parkinson's disease, Alzheimer's disease, Huntington's chorea, or amyotrophic lateral sclerosis;
- (vi) lesions associated with nutritional diseases or disorders, in which a portion of the
- 30 nervous system is destroyed or injured by a nutritional disorder or disorder of metabolism including but not limited to, vitamin B12 deficiency, folic acid deficiency, Wernicke
- 35 disease, tobacco-alcohol amblyopia, Marchiafava-Bignami disease (primary

- degeneration of the corpus callosum), and alcoholic cerebellar degeneration;
- 5 (vii) neurological lesions associated with systemic diseases including but not limited to diabetes (diabetic neuropathy, Bell's palsy), systemic lupus erythematosus, carcinoma, or sarcoidosis;
- (viii) lesions caused by toxic substances including alcohol, lead, or particular neurotoxins; and
- 10 (ix) demyelinated lesions in which a portion of the nervous system is destroyed or injured by a demyelinating disease including but not limited to multiple sclerosis, human immunodeficiency virus-associated myelopathy,
- 15 transverse myelopathy or various etiologies, progressive multifocal leukoencephalopathy, and central pontine myelinolysis.

Therapeutics which are useful according to the invention for treatment of a nervous system disorder may be

20 selected by testing for biological activity in promoting the survival or differentiation of neurons (see also Section 5.8). For example, and not by way of limitation, Therapeutics which elicit any of the following effects may be useful according to the invention:

- 25 (i) increased survival time of neurons in culture;
- (ii) increased sprouting of neurons in culture or *in vivo*;
- (iii) increased production of a neuron-associated molecule in culture or *in vivo*, e.g., choline acetyltransferase or acetylcholinesterase with
- 30 respect to motor neurons; or
- (iv) decreased symptoms of neuron dysfunction *in vivo*.

Such effects may be measured by any method known in the art.

35 In preferred, non-limiting embodiments, increased survival of neurons may be measured by the method set forth in Arakawa et al. (1990, J. Neurosci. 10:3507-3515); increased sprouting of

neurons may be detected by methods set forth in Pestronk et al. (1980, Exp. Neurol. 70:65-82) or Brown et al. (1981, Ann. Rev. Neurosci. 4:17-42); increased production of neuron-associated molecules may be measured by bioassay, enzymatic  
5 assay, antibody binding, Northern blot assay, etc., depending on the molecule to be measured; and motor neuron dysfunction may be measured by assessing the physical manifestation of motor neuron disorder, e.g., weakness, motor neuron conduction velocity, or functional disability.

10 In a specific embodiments, motor neuron disorders that may be treated according to the invention include but are not limited to disorders such as infarction, infection, exposure to toxin, trauma, surgical damage, degenerative disease or malignancy that may affect motor neurons as well  
15 as other components of the nervous system, as well as disorders that selectively affect neurons such as amyotrophic lateral sclerosis, and including but not limited to progressive spinal muscular atrophy, progressive bulbar palsy, primary lateral sclerosis, infantile and juvenile  
20 muscular atrophy, progressive bulbar paralysis of childhood (Fazio-Londe syndrome), poliomyelitis and the post polio syndrome, and Hereditary Motorsensory Neuropathy (Charcot-Marie-Tooth Disease).

#### 25 5.8.3. TISSUE REPAIR AND REGENERATION

In another embodiment of the invention, a  
Therapeutic of the invention is used for promotion of tissue  
regeneration and repair, including but not limited to  
treatment of benign dysproliferative disorders. Specific  
30 embodiments are directed to treatment of cirrhosis of the liver (a condition in which scarring has overtaken normal liver regeneration processes), treatment of keloid  
(hypertrophic scar) formation (disfiguring of the skin in  
which the scarring process interferes with normal renewal),  
35 psoriasis (a common skin condition characterized by excessive proliferation of the skin and delay in proper cell fate determination), and baldness (a condition in which terminally

differentiated hair follicles (a tissue rich in Notch) fail to function properly). In another embodiment, a Therapeutic of the invention is used to treat degenerative or traumatic disorders of the sensory epithelium of the inner ear.

5

## 5.9. PROPHYLACTIC USES

### 5.9.1. MALIGNANCIES

The Therapeutics of the invention can be administered to prevent progression to a neoplastic or malignant state, including but not limited to those disorders listed in Table 1. Such administration is indicated where the Therapeutic is shown in assays, as described supra, to have utility for treatment or prevention of such disorder. Such prophylactic use is indicated in conditions known or suspected of preceding progression to neoplasia or cancer, in particular, where non-neoplastic cell growth consisting of hyperplasia, metaplasia, or most particularly, dysplasia has occurred (for review of such abnormal growth conditions, see Robbins and Angell, 1976, *Basic Pathology*, 2d Ed., W.B. Saunders Co., Philadelphia, pp. 68-79.) Hyperplasia is a form of controlled cell proliferation involving an increase in cell number in a tissue or organ, without significant alteration in structure or function. As but one example, endometrial hyperplasia often precedes endometrial cancer. Metaplasia is a form of controlled cell growth in which one type of adult or fully differentiated cell substitutes for another type of adult cell. Metaplasia can occur in epithelial or connective tissue cells. Atypical metaplasia involves a somewhat disorderly metaplastic epithelium. Dysplasia is frequently a forerunner of cancer, and is found mainly in the epithelia; it is the most disorderly form of non-neoplastic cell growth, involving a loss in individual cell uniformity and in the architectural orientation of cells. Dysplastic cells often have abnormally large, deeply stained nuclei, and exhibit pleomorphism. Dysplasia characteristically occurs where there exists chronic

irritation or inflammation, and is often found in the cervix, respiratory passages, oral cavity, and gall bladder.

Alternatively or in addition to the presence of abnormal cell growth characterized as hyperplasia, 5 metaplasia, or dysplasia, the presence of one or more characteristics of a transformed phenotype, or of a malignant phenotype, displayed *in vivo* or displayed *in vitro* by a cell sample from a patient, can indicate the desirability of prophylactic/therapeutic administration of a Therapeutic of 10 the invention. As mentioned *supra*, such characteristics of a transformed phenotype include morphology changes, looser substratum attachment, loss of contact inhibition, loss of anchorage dependence, protease release, increased sugar transport, decreased serum requirement, expression of fetal 15 antigens, disappearance of the 250,000 dalton cell surface protein, etc. (see also *id.*, at pp. 84-90 for characteristics associated with a transformed or malignant phenotype).

In a specific embodiment, leukoplakia, a benign-appearing hyperplastic or dysplastic lesion of the 20 epithelium, or Bowen's disease, a carcinoma *in situ*, are pre-neoplastic lesions indicative of the desirability of prophylactic intervention.

In another embodiment, fibrocystic disease (cystic hyperplasia, mammary dysplasia, particularly adenosis (benign 25 epithelial hyperplasia)) is indicative of the desirability of prophylactic intervention.

In other embodiments, a patient which exhibits one or more of the following predisposing factors for malignancy is treated by administration of an effective amount of a 30 Therapeutic: a chromosomal translocation associated with a malignancy (e.g., the Philadelphia chromosome for chronic myelogenous leukemia, t(14;18) for follicular lymphoma, etc.), familial polyposis or Gardner's syndrome (possible forerunners of colon cancer), benign monoclonal gammopathy (a 35 possible forerunner of multiple myeloma), and a first degree kinship with persons having a cancer or precancerous disease showing a Mendelian (genetic) inheritance pattern (e.g.,



familial polyposis of the colon, Gardner's syndrome, hereditary exostosis, polyendocrine adenomatosis, medullary thyroid carcinoma with amyloid production and pheochromocytoma, Peutz-Jeghers syndrome, neurofibromatosis of Von Recklinghausen, retinoblastoma, carotid body tumor, cutaneous melanocarcinoma, intraocular melanocarcinoma, xeroderma pigmentosum, ataxia telangiectasia, Chediak-Higashi syndrome, albinism, Fanconi's aplastic anemia, and Bloom's syndrome; see Robbins and Angell, 1976, *Basic Pathology*, 2d Ed., W.B. Saunders Co., Philadelphia, pp. 112-113) etc.)

In another specific embodiment, an Antagonist Therapeutic of the invention is administered to a human patient to prevent progression to breast, colon, or cervical cancer.

15

#### 5.9.2. OTHER DISORDERS

In other embodiments, a Therapeutic of the invention can be administered to prevent a nervous system disorder described in Section 5.8.2, or other disorder (e.g., liver cirrhosis, psoriasis, keloids, baldness) described in Section 5.8.3.

#### 5.10. DEMONSTRATION OF THERAPEUTIC OR PROPHYLACTIC UTILITY

The Therapeutics of the invention can be tested in vivo for the desired therapeutic or prophylactic activity. For example, such compounds can be tested in suitable animal model systems prior to testing in humans, including but not limited to rats, mice, chicken, cows, monkeys, rabbits, etc. For in vivo testing, prior to administration to humans, any animal model system known in the art may be used.

#### 5.11. ANTISENSE REGULATION OF SERRATE EXPRESSION

The present invention provides the therapeutic or prophylactic use of nucleic acids of at least six or of at least ten nucleotides that are antisense to a gene or cDNA encoding a vertebrate Serrate or a portion thereof.

"Antisense" as used herein refers to a nucleic acid capable of hybridizing to a portion of a vertebrate Serrate RNA (preferably mRNA) by virtue of some sequence complementarity. Such antisense nucleic acids have utility as Antagonist  
5 Therapeutics of the invention, and can be used in the treatment or prevention of disorders as described *supra* in Section 5.8 and its subsections.

The antisense nucleic acids of the invention can be oligonucleotides that are double-stranded or single-stranded,  
10 RNA or DNA or a modification or derivative thereof, which can be directly administered to a cell, or which can be produced intracellularly by transcription of exogenous, introduced sequences.

In a specific embodiment, the Serrate antisense  
15 nucleic acids provided by the instant invention can be used for the treatment of tumors or other disorders, the cells of which tumor type or disorder can be demonstrated (*in vitro* or *in vivo*) to express a Serrate gene or a Notch gene. Such demonstration can be by detection of RNA or of protein.

20 The invention further provides pharmaceutical compositions comprising an effective amount of the Serrate antisense nucleic acids of the invention in a pharmaceutically acceptable carrier, as described *infra* in Section 5.12. Methods for treatment and prevention of  
25 disorders (such as those described in Sections 5.8 and 5.9) comprising administering the pharmaceutical compositions of the invention are also provided.

In another embodiment, the invention is directed to methods for inhibiting the expression of a Serrate nucleic  
30 acid sequence in a prokaryotic or eukaryotic cell comprising providing the cell with an effective amount of a composition comprising an antisense vertebrate Serrate nucleic acid of the invention.

Serrate antisense nucleic acids and their uses are  
35 described in detail below.

### 5.11.1. VERTEBRATE SERRATE ANTISENSE NUCLEIC ACIDS

The vertebrate Serrate antisense nucleic acids are of at least six nucleotides and are preferably oligonucleotides (ranging preferably from 10 to about 50 oligonucleotides). In specific aspects, the oligonucleotide contains at least 10 nucleotides, at least 15 nucleotides, at least 100 nucleotides, or at least 200 nucleotides antisense to a Serrate gene. The oligonucleotides can be DNA or RNA or chimeric mixtures or derivatives or modified versions thereof, single-stranded or double-stranded. The oligonucleotide can be modified at the base moiety, sugar moiety, or phosphate backbone. The oligonucleotide may include other appending groups such as peptides, or agents facilitating transport across the cell membrane (see, e.g., Letsinger et al., 1989, Proc. Natl. Acad. Sci. U.S.A. 86:6553-6556; Lemaitre et al., 1987, Proc. Natl. Acad. Sci. 84:648-652; PCT Publication No. WO 88/09810, published December 15, 1988) or blood-brain barrier (see, e.g., PCT Publication No. WO 89/10134, published April 25, 1988), hybridization-triggered cleavage agents (see, e.g., Krol et al., 1988, BioTechniques 6:958-976) or intercalating agents (see, e.g., Zon, 1988, Pharm. Res. 5:539-549).

In a preferred aspect of the invention, a vertebrate Serrate antisense oligonucleotide is provided, preferably of single-stranded DNA. In a most preferred aspect, such an oligonucleotide comprises a sequence antisense to the sequence encoding an SH3 binding domain or a Notch-binding domain of Serrate, most preferably, of a human Serrate homolog. The oligonucleotide may be modified at any position on its structure with substituents generally known in the art.

The Serrate antisense oligonucleotide may comprise at least one modified base moiety which is selected from the group including but not limited to 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xantine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine,

5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, 10 queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine.

15 In another embodiment, the oligonucleotide comprises at least one modified sugar moiety selected from the group including but not limited to arabinose, 2-fluoroarabinose, xylulose, and hexose.

In yet another embodiment, the oligonucleotide 20 comprises at least one modified phosphate backbone selected from the group consisting of a phosphorothioate, a phosphorodithioate, a phosphoramidothioate, a phosphoramidate, a phosphordiamidate, a methylphosphonate, an alkyl phosphotriester, and a formacetal or analog thereof.

25 In yet another embodiment, the oligonucleotide is an  $\alpha$ -anomeric oligonucleotide. An  $\alpha$ -anomeric oligonucleotide forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual  $\beta$ -units, the strands run parallel to each other (Gautier et al., 1987, Nucl. Acids 30 Res. 15:6625-6641).

The oligonucleotide may be conjugated to another molecule, e.g., a peptide, hybridization triggered cross-linking agent, transport agent, hybridization-triggered cleavage agent, etc.

35 Oligonucleotides of the invention may be synthesized by standard methods known in the art, e.g. by use of an automated DNA synthesizer (such as are commercially

available from Biosearch, Applied Biosystems, etc.). As examples, phosphorothioate oligonucleotides may be synthesized by the method of Stein et al. (1988, Nucl. Acids Res. 16:3209), methylphosphonate oligonucleotides can be  
5 prepared by use of controlled pore glass polymer supports (Sarin et al., 1988, Proc. Natl. Acad. Sci. U.S.A. 85:7448-7451), etc.

In a specific embodiment, the Serrate antisense oligonucleotide comprises catalytic RNA, or a ribozyme (see,  
10 e.g., PCT International Publication WO 90/11364, published October 4, 1990; Sarver et al., 1990, Science 247:1222-1225). In another embodiment, the oligonucleotide is a 2'-O-methylribonucleotide (Inoue et al., 1987, Nucl. Acids Res. 15:6131-6148), or a chimeric RNA-DNA analogue (Inoue et al.,  
15 1987, FEBS Lett. 215:327-330).

In an alternative embodiment, the Serrate antisense nucleic acid of the invention is produced intracellularly by transcription from an exogenous sequence. For example, a vector can be introduced in vivo such that it is taken up by  
20 a cell, within which cell the vector or a portion thereof is transcribed, producing an antisense nucleic acid (RNA) of the invention. Such a vector would contain a sequence encoding the Serrate antisense nucleic acid. Such a vector can remain episomal or become chromosomally integrated, as long as it  
25 can be transcribed to produce the desired antisense RNA. Such vectors can be constructed by recombinant DNA technology methods standard in the art. Vectors can be plasmid, viral, or others known in the art, used for replication and expression in mammalian cells. Expression of the sequence  
30 encoding the Serrate antisense RNA can be by any promoter known in the art to act in mammalian, preferably human, cells. Such promoters can be inducible or constitutive. Such promoters include but are not limited to: the SV40 early promoter region (Bernoist and Chambon, 1981, Nature 290:304-  
35 310), the promoter contained in the 3' long terminal repeat of Rous sarcoma virus (Yamamoto et al., 1980, Cell 22:787-797), the herpes thymidine kinase promoter (Wagner et al.,

1981, Proc. Natl. Acad. Sci. U.S.A. 78:1441-1445), the regulatory sequences of the metallothionein gene (Brinster et al., 1982, Nature 296:39-42), etc.

The antisense nucleic acids of the invention  
5 comprise a sequence complementary to at least a portion of an RNA transcript specific to a vertebrate Serrate gene, preferably a human Serrate gene. However, absolute complementarity, although preferred, is not required. A sequence "complementary to at least a portion of an RNA," as  
10 referred to herein, means a sequence having sufficient complementarity to be able to hybridize with the RNA, forming a stable duplex; in the case of double-stranded Serrate antisense nucleic acids, a single strand of the duplex DNA may thus be tested, or triplex formation may be assayed. The  
15 ability to hybridize will depend on both the degree of complementarity and the length of the antisense nucleic acid. Generally, the longer the hybridizing nucleic acid, the more base mismatches with a Serrate RNA it may contain and still form a stable duplex (or triplex, as the case may be). One  
20 skilled in the art can ascertain a tolerable degree of mismatch by use of standard procedures to determine the melting point of the hybridized complex.

25 5.11.2. THERAPEUTIC UTILITY OF VERTEBRATE  
SERRATE ANTISENSE NUCLEIC ACIDS

The vertebrate Serrate antisense nucleic acids can be used to treat (or prevent) malignancies or other disorders, of a cell type which has been shown to express Serrate or Notch. In specific embodiments, the malignancy is  
30 cervical, breast, or colon cancer, or squamous adenocarcinoma. Malignant, neoplastic, and pre-neoplastic cells which can be tested for such expression include but are not limited to those described supra in Sections 5.8.1 and 5.9.1. In a preferred embodiment, a single-stranded DNA  
35 antisense Serrate oligonucleotide is used.

Malignant (particularly, tumor) cell types which express Serrate or Notch RNA can be identified by various

methods known in the art. Such methods include but are not limited to hybridization with a *Serrate* or *Notch*-specific nucleic acid (e.g. by Northern hybridization, dot blot hybridization, *in situ* hybridization), observing the ability of RNA from the cell type to be translated *in vitro* into *Notch* or *Serrate*, immunoassay, etc. In a preferred aspect, primary tumor tissue from a patient can be assayed for *Notch* or *Serrate* expression prior to treatment, e.g., by immunocytochemistry or *in situ* hybridization.

10           Pharmaceutical compositions of the invention (see Section 5.12), comprising an effective amount of a vertebrate *Serrate* antisense nucleic acid in a pharmaceutically acceptable carrier, can be administered to a patient having a malignancy which is of a type that expresses *Notch* or *Serrate* RNA or protein.

          The amount of *Serrate* antisense nucleic acid which will be effective in the treatment of a particular disorder or condition will depend on the nature of the disorder or condition, and can be determined by standard clinical techniques. Where possible, it is desirable to determine the antisense cytotoxicity of the tumor type to be treated *in vitro*, and then in useful animal model systems prior to testing and use in humans.

          In a specific embodiment, pharmaceutical compositions comprising vertebrate *Serrate* antisense nucleic acids are administered via liposomes, microparticles, or microcapsules. In various embodiments of the invention, it may be useful to use such compositions to achieve sustained release of the *Serrate* antisense nucleic acids. In a specific embodiment, it may be desirable to utilize liposomes targeted via antibodies to specific identifiable tumor antigens (Leonetti et al., 1990, Proc. Natl. Acad. Sci. U.S.A. 87:2448-2451; Renneisen et al., 1990, J. Biol. Chem. 265:16337-16342).

35

5.12. THERAPEUTIC/PROPHYLACTIC  
ADMINISTRATION AND COMPOSITIONS

The invention provides methods of treatment (and prophylaxis) by administration to a subject of an effective amount of a Therapeutic of the invention. In a preferred aspect, the Therapeutic is substantially purified. The subject is preferably an animal, including but not limited to animals such as cows, pigs, chickens, etc., and is preferably a mammal, and most preferably human.

Various delivery systems are known and can be used to administer a Therapeutic of the invention, e.g., encapsulation in liposomes, microparticles, microcapsules, expression by recombinant cells, receptor-mediated endocytosis (see, e.g., Wu and Wu, 1987, J. Biol. Chem. 262:4429-4432), construction of a Therapeutic nucleic acid as part of a retroviral or other vector, etc. Methods of introduction include but are not limited to intradermal, intramuscular, intraperitoneal, intravenous, subcutaneous, intranasal, epidural, and oral routes. The compounds may be administered by any convenient route, for example by infusion or bolus injection, by absorption through epithelial or mucocutaneous linings (e.g., oral mucosa, rectal and intestinal mucosa, etc.) and may be administered together with other biologically active agents. Administration can be systemic or local. In addition, it may be desirable to introduce the pharmaceutical compositions of the invention into the central nervous system by any suitable route, including intraventricular and intrathecal injection; intraventricular injection may be facilitated by an intraventricular catheter, for example, attached to a reservoir, such as an Ommaya reservoir. Pulmonary administration can also be employed, e.g., by use of an inhaler or nebulizer, and formulation with an aerosolizing agent.

In a specific embodiment, it may be desirable to administer the pharmaceutical compositions of the invention locally to the area in need of treatment; this may be



achieved by, for example, and not by way of limitation, local infusion during surgery, topical application, e.g., in conjunction with a wound dressing after surgery, by injection, by means of a catheter, by means of a suppository, 5 or by means of an implant, said implant being of a porous, non-porous, or gelatinous material, including membranes, such as sialastic membranes, or fibers. In one embodiment, administration can be by direct injection at the site (or former site) of a malignant tumor or neoplastic or pre- 10 neoplastic tissue.

In another embodiment, the Therapeutic can be delivered in a vesicle, in particular a liposome (see Langer, Science 249:1527-1533 (1990); Treat et al., in Liposomes in the Therapy of Infectious Disease and Cancer, Lopez-Berestein 15 and Fidler (eds.), Liss, New York, pp. 353-365 (1989); Lopez-Berestein, *ibid.*, pp. 317-327; see generally *ibid.*)

In yet another embodiment, the Therapeutic can be delivered in a controlled release system. In one embodiment, a pump may be used (see Langer, *supra*; Sefton, CRC Crit. Ref. 20 Biomed. Eng. 14:201 (1987); Buchwald et al., Surgery 88:507 (1980); Saudek et al., N. Engl. J. Med. 321:574 (1989)). In another embodiment, polymeric materials can be used (see Medical Applications of Controlled Release, Langer and Wise (eds.), CRC Pres., Boca Raton, Florida (1974); Controlled 25 Drug Bioavailability, Drug Product Design and Performance, Smolen and Ball (eds.), Wiley, New York (1984); Ranger and Peppas, J. Macromol. Sci. Rev. Macromol. Chem. 23:61 (1983); see also Levy et al., Science 228:190 (1985); During et al., Ann. Neurol. 25:351 (1989); Howard et al., J. Neurosurg. 30 71:105 (1989)). In yet another embodiment, a controlled release system can be placed in proximity of the therapeutic target, i.e., the brain, thus requiring only a fraction of the systemic dose (see, e.g., Goodson, in Medical Applications of Controlled Release, *supra*, vol. 2, pp. 35 115-138 (1984)).

Other controlled release systems are discussed in the review by Langer (Science 249:1527-1533 (1990)).

In a specific embodiment where the Therapeutic is a nucleic acid encoding a protein Therapeutic, the nucleic acid can be administered *in vivo* to promote expression of its encoded protein, by constructing it as part of an appropriate nucleic acid expression vector and administering it so that it becomes intracellular, e.g., by use of a retroviral vector (see U.S. Patent No. 4,980,286), or by direct injection, or by use of microparticle bombardment (e.g., a gene gun; Biolistic, Dupont), or coating with lipids or cell-surface receptors or transfecting agents, or by administering it in linkage to a homeobox-like peptide which is known to enter the nucleus (see e.g., Joliot et al., 1991, Proc. Natl. Acad. Sci. USA 88:1864-1868), etc. Alternatively, a nucleic acid Therapeutic can be introduced intracellularly and incorporated within host cell DNA for expression, by homologous recombination.

In specific embodiments directed to treatment or prevention of particular disorders, preferably the following forms of administration are used:

20	<u>Disorder</u>	<u>Preferred Forms of Administration</u>
	Cervical cancer	Topical
	Gastrointestinal cancer	Oral; intravenous
	Lung cancer	Inhaled; intravenous
	Leukemia	Intravenous; extracorporeal
25	Metastatic carcinomas	Intravenous; oral
	Brain cancer	Targeted; intravenous; intrathecal
	Liver cirrhosis	Oral; intravenous
	Psoriasis	Topical
	Keloids	Topical
30	Baldness	Topical
	Spinal cord injury	Targeted; intravenous; intrathecal
	Parkinson's disease	Targeted; intravenous; intrathecal
	Motor neuron disease	Targeted; intravenous; intrathecal
35	Alzheimer's disease	Targeted; intravenous; intrathecal

The present invention also provides pharmaceutical compositions. Such compositions comprise a therapeutically effective amount of a Therapeutic, and a pharmaceutically acceptable carrier. In a specific embodiment, the term

5 "pharmaceutically acceptable" means approved by a regulatory agency of the Federal or a state government or listed in the U.S. Pharmacopeia or other generally recognized pharmacopeia for use in animals, and more particularly in humans. The term "carrier" refers to a diluent, adjuvant, excipient, or

10 vehicle with which the therapeutic is administered. Such pharmaceutical carriers can be sterile liquids, such as water and oils, including those of petroleum, animal, vegetable or synthetic origin, such as peanut oil, soybean oil, mineral oil, sesame oil and the like. Water is a preferred carrier

15 when the pharmaceutical composition is administered intravenously. Saline solutions and aqueous dextrose and glycerol solutions can also be employed as liquid carriers, particularly for injectable solutions. Suitable pharmaceutical excipients include starch, glucose, lactose,

20 sucrose, gelatin, malt, rice, flour, chalk, silica gel, sodium stearate, glycerol monostearate, talc, sodium chloride, dried skim milk, glycerol, propylene, glycol, water, ethanol and the like. The composition, if desired, can also contain minor amounts of wetting or emulsifying

25 agents, or pH buffering agents. These compositions can take the form of solutions, suspensions, emulsion, tablets, pills, capsules, powders, sustained-release formulations and the like. The composition can be formulated as a suppository, with traditional binders and carriers such as triglycerides.

30 Oral formulation can include standard carriers such as pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, etc. Examples of suitable pharmaceutical carriers are described in "Remington's Pharmaceutical Sciences" by E.W.

35 Martin. Such compositions will contain a therapeutically effective amount of the Therapeutic, preferably in purified form, together with a suitable amount of carrier so as to

provide the form for proper administration to the patient. The formulation should suit the mode of administration.

In a preferred embodiment, the composition is formulated in accordance with routine procedures as a pharmaceutical composition adapted for intravenous administration to human beings. Typically, compositions for intravenous administration are solutions in sterile isotonic aqueous buffer. Where necessary, the composition may also include a solubilizing agent and a local anesthetic such as lignocaine to ease pain at the site of the injection. Generally, the ingredients are supplied either separately or mixed together in unit dosage form, for example, as a dry lyophilized powder or water free concentrate in a hermetically sealed container such as an ampoule or sachette indicating the quantity of active agent. Where the composition is to be administered by infusion, it can be dispensed with an infusion bottle containing sterile pharmaceutical grade water or saline. Where the composition is administered by injection, an ampoule of sterile water for injection or saline can be provided so that the ingredients may be mixed prior to administration.

The Therapeutics of the invention can be formulated as neutral or salt forms. Pharmaceutically acceptable salts include those formed with free amino groups such as those derived from hydrochloric, phosphoric, acetic, oxalic, tartaric acids, etc., and those formed with free carboxyl groups such as those derived from sodium, potassium, ammonium, calcium, ferric hydroxides, isopropylamine, triethylamine, 2-ethylamino ethanol, histidine, procaine, etc.

The amount of the Therapeutic of the invention which will be effective in the treatment of a particular disorder or condition will depend on the nature of the disorder or condition, and can be determined by standard clinical techniques. In addition, *in vitro* assays may optionally be employed to help identify optimal dosage ranges. The precise dose to be employed in the formulation

will also depend on the route of administration, and the seriousness of the disease or disorder, and should be decided according to the judgment of the practitioner and each patient's circumstances. However, suitable dosage ranges for intravenous administration are generally about 20-500 micrograms of active compound per kilogram body weight. Suitable dosage ranges for intranasal administration are generally about 0.01 pg/kg body weight to 1 mg/kg body weight. Effective doses may be extrapolated from dose-response curves derived from *in vitro* or animal model test systems.

Suppositories generally contain active ingredient in the range of 0.5% to 10% by weight; oral formulations preferably contain 10% to 95% active ingredient.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Optionally associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration.

#### 5.13. DIAGNOSTIC UTILITY

Vertebrate Serrate proteins, analogues, derivatives, and subsequences thereof, vertebrate Serrate nucleic acids (and sequences complementary thereto), anti-vertebrate Serrate antibodies, have uses in diagnostics. Such molecules can be used in assays, such as immunoassays, to detect, prognose, diagnose, or monitor various conditions, diseases, and disorders affecting Serrate expression, or monitor the treatment thereof. In particular, such an immunoassay is carried out by a method comprising contacting a sample derived from a patient with an anti-Serrate antibody under conditions such that immunospecific binding can occur, and detecting or measuring the amount of any immunospecific

binding by the antibody. In a specific aspect, such binding of antibody, in tissue sections, preferably in conjunction with binding of anti-Notch antibody can be used to detect aberrant Notch and/or Serrate localization or aberrant levels of Notch-Serrate colocalization in a disease state. In a specific embodiment, antibody to Serrate can be used to assay in a patient tissue or serum sample for the presence of Serrate where an aberrant level of Serrate is an indication of a diseased condition. Aberrant levels of Serrate binding ability in an endogenous Notch protein, or aberrant levels of binding ability to Notch (or other Serrate ligand) in an endogenous Serrate protein may be indicative of a disorder of cell fate (e.g., cancer, etc.) By "aberrant levels," is meant increased or decreased levels relative to that present, or a standard level representing that present, in an analogous sample from a portion of the body or from a subject not having the disorder.

The immunoassays which can be used include but are not limited to competitive and non-competitive assay systems using techniques such as western blots, radioimmunoassays, ELISA (enzyme linked immunosorbent assay), "sandwich" immunoassays, immunoprecipitation assays, precipitin reactions, gel diffusion precipitin reactions, immunodiffusion assays, agglutination assays, complement-fixation assays, immunoradiometric assays, fluorescent immunoassays, protein A immunoassays, to name but a few.

Vertebrate Serrate genes and related nucleic acid sequences and subsequences, including complementary sequences, and other toporythmic gene sequences, can also be used in hybridization assays. Vertebrate Serrate nucleic acid sequences, or subsequences thereof comprising about at least 8 nucleotides, can be used as hybridization probes. Hybridization assays can be used to detect, prognose, diagnose, or monitor conditions, disorders, or disease states associated with aberrant changes in Serrate expression and/or activity as described supra. In particular, such a hybridization assay is carried out by a method comprising

contacting a sample containing nucleic acid with a nucleic acid probe capable of hybridizing to Serrate DNA or RNA, under conditions such that hybridization can occur, and detecting or measuring any resulting hybridization.

- 5                    Additionally, since Serrate binds to Notch, vertebrate Serrate or a binding portion thereof can be used to assay for the presence and/or amounts of Notch in a sample, e.g., in screening for malignancies which exhibit increased Notch expression such as colon and cervical  
10 cancers.

#### 6.    ISOLATION AND CHARACTERIZATION OF A MOUSE SERRATE HOMOLOG

A mouse Serrate homolog, termed M-Serrate-1, was  
15 isolated as follows:

##### **Mouse Serrate-1 gene**

Tissue origin: 10.5-day mouse embryonic RNA

Isolation method:

- a) random primed cDNA against above RNA  
b) PCR of above cDNA using  
20 PCR primer 1: CGI(C/T)TTTGC(C/T)TIAA(A/G)(G/C)AITA(C/T)CA  
(SEQ ID NO: 9) {encoding RLCKK(H/E)YQ (SEQ ID NO:10)}:  
PCR primer 2: TCIATGCAIGTICCCIC(A/G)TT (SEQ ID NO:11)  
{encoding NGGTCID (SEQ ID NO:12)}  
25 Amplification conditions: 50 ng cDNA, 1 µg each primer,  
0.2 mM dNTP's, 1.8 U Taq (Perkin-Elmer) in 50 µl of supplied  
buffer, 40 cycles of: 94°C/30 sec, 45°C/2 min, 72°C/1 min  
extended by 2 sec each cycle.

- 30 Yielded a 1.8 kb fragment which was sequenced at both ends  
and identified as corresponding to C-Serrate-1

Partial DNA sequence of M-Serrate-1:

From 5' end:

- 35 GTCCCGCGTCACTGCCGGGGGACCCTGCAGCTTCGGCTCAGGGTCTACGCCTGTCATCGGG  
GGTAACACCTTCAATCTCAAGGCCAGCCGTGGCAACGACCGTAATCGCATCGTACTGCCTT  
TCAGTTTACCTGGCCGAGGTCCTACACTTTGCTGGTGGAG (SEQ ID NO:13)

Protein translation of above:

SRVTAGGPCSFSGSGSTPVIGGNTFNLKASRGNDRNRIVLPFSFTWPRSYTLLVE  
(SEQ ID NO:14) (corresponds to amino-terminal sequence  
upstream of the DSL domain)

5

From 3' end (but coding strand)

TCTTCTAACGTCTGTGGTCCCCATGGCAAGTGCAAGAGCCAGTCGGCAGGCAAATTCACCT  
GTGACTGTAACAAAGGCTTCACCGGCACCTACTGCCATGAAAATATCAACGACTGCGAGAG  
CAACCCCTGTAAA (SEQ ID NO:15)

10 Protein translation of above:

SSNVCGPHGKCKSQSAGKFTCDCNKGFTGTYPHENINDCESNPCK (SEQ ID NO:16)  
(within tandemly arranged EGF-like repeats)

Expression pattern: The expression pattern was determined to  
15 be the same as that observed for C-Serrate-1 (chicken  
Serrate) (see Section 11 *infra*), including expression in the  
developing central nervous system, peripheral nervous system,  
limb, kidney, lens, and vascular system.

20

#### 7. ISOLATION AND CHARACTERIZATION OF A XENOPUS SERRATE HOMOLOG

A *Xenopus* Serrate homolog, termed *Xenopus* Serrate-1  
was isolated as follows:

***Xenopus* Serrate-1 gene**

25 Tissue origin: neurula-stage embryonic RNA

Isolation method:

a) random primed cDNA against above RNA

b) PCR using:

Primer 1: CGI(C/T)TTTGC(C/T)TIAA(A/G)(G/C)AITA(C/T)CA  
(SEQ ID NO:9) {encoding RLCK(H/E)YQ (SEQ ID NO:10)}:

30 PCR primer 2: TCIATGCAIGTICCC(A/G)TT (SEQ ID NO:11)  
{encoding NGGTCID (SEQ ID NO:12)}

Amplification conditions: 50 ng cDNA, 1 µg each primer,  
0.2 mM dNTP's, 1.8 U Taq (Perkin-Elmer) in 50 µl of supplied  
35 buffer. 40 cycles of: 94°C/30 sec, 45°C/2 min, 72°C/1 min  
extended by 2 sec each cycle.



Yielded a -700 bp fragment which was partially sequenced to confirm its relationship to C-Serrate-1.

#### 8. ISOLATION AND CHARACTERIZATION OF A CHICK SERRATE HOMOLOG

5 In the example herein, we report the cloning and sequence of a chick Serrate homolog, C-Serrate, and of fragments of two chick Notch homologs, C-Notch-1 and C-Notch-2, together with their expression patterns during  
10 early embryogenesis. The patterns of transcription of C-Serrate overlaps with that of C-Notch-1 in many regions of the embryo, suggesting that C-Notch-1, like Notch in *Drosophila*, is a receptor for Serrate. In particular, Notch and Serrate are expressed in the neurogenic regions of the  
15 developing central and peripheral nervous system.

Our data show that Serrate, a known ligand of Notch, has been conserved from arthropods to chordates. The overlapping expression patterns suggest conservation of its functional relationship with Notch and imply that development  
20 of the chick and in particular of its central nervous system involves the interaction of C-Notch-1 with Serrate at several specific locations.

#### Materials and Methods

##### 25 Embryos

White Leghorn chicken eggs were obtained from University Park Farm and incubated at 38°C. Embryos were staged according to Hamburger and Hamilton (1951, J. Exp. Zool. 88:49-92).

##### 30 Cloning of chicken homologs of Notch

Approximately 1000 base pair PCR fragments of the chicken Notch 1 and Notch 2 genes were amplified from otic explant RNA (see below) using degenerate primers and PCR  
35 conditions as outlined in Lardelli and Lendahl (1993, Exp. Cell Res. 204:364-372). The PCR fragment was subcloned into Bluescript KS-, sequenced and used as a template for making a

DIG antisense RNA probe (RNA Transcription Kit, Stratagene; DIG RNA labelling mix, Boehringer Mannheim).

*Cloning of a chicken homologue of Drosophila Serrate*

5           Otic explants were dissected from embryos of stages 8 to 13. Each otic explant consisted of the two otic cups, a short section of intervening hindbrain and pharynx and the associated head ectoderm and mesenchyme. RNA was extracted using a modification of standard protocols (Sambrook et al., 10 1989, in *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York) and polyA<sup>+</sup> mRNA was isolated from total RNA using the PolyATtract mRNA Isolation System (Promega). First strand cDNA was synthesized using the SuperScript Preamplification 15 System (Gibco).

          PCR and degenerate primers were used to amplify a fragment of a chicken gene homologous to the *Drosophila* gene *Serrate* from the otic explant cDNA. The primers were designed to recognize peptide motifs found in both the fly 20 *Delta* and *Serrate* proteins:

- 1) primer 1, 5-CGI(T/C)TITGC(T/C)TIAA(G/A)(G/C)AITA(C/T)CA-3' (SEQ ID NO:17), corresponds to the motif RLCLK(E/H)YQ (SEQ ID NO:18) located at the amino-terminus of the fly *Delta* and *Serrate* proteins.
- 25 2) primer 2, 5'-TCIATGCAIGTICCCIC(A/G)TT-3' (SEQ ID NO:11), corresponds to the motif NGGTCID (SEQ ID NO:12) found in several of the EGF-like repeats. The PCR conditions were as follows: 35 cycles of 94°C for 1 minute, 45°C for 1.5 minutes and 72°C for 2 minutes; followed by a final extension step of 30 72°C for 10 minutes. A PCR product of approximately 900 base pairs in length was purified, subcloned into Bluescript KS- (Stratagene) and its DNA sequence partially determined to confirm that it was a likely *Serrate* homolog. It was then used to recover larger cDNA clones by screening two cDNA 35 libraries:

- 1) a stage 8-13 otic explant random primed cDNA library

2) a stage 17 chick spinal cord oligo dT primed cDNA library overlapping cDNAs were isolated, and two (termed 9 and 3A.1) that together cover almost the entire coding region of the gene were subcloned into Bluescript KS-. DNA sequence was  
5 determined from nested deletion series generated using the double-stranded Nested Deletion Kit (Pharmacia) and Sanger dideoxy chain termination method with the Sequenase enzyme (US Biochemical Corporation). Sequences were aligned and analyzed using Geneworks 2.3 and Intelligenetics. Homology  
10 searches were done using the program Sharq.

To obtain the most 5' end of the open reading frame, a number of other PCR based strategies were used including the screening of a number of other libraries (cDNA and genomic) using the method of Lardelli et al. (1994,  
15 Mechanisms of Development 46:123-136).

#### *In situ hybridization*

Patterns of gene transcription were determined by *in situ* hybridization using DIG-labeled RNA probes and:  
20 1) a high-stringency wholemount *in situ* hybridization protocol, and  
2) *in situ* hybridization on cryostat sections based on the protocol of Strähle et al. (1994, Trends in Genet. 10:7).

#### **25 Results**

To obtain insight into the likely role of chick Serrate in the vertebrate embryo, we examined its expression in relation to that of chick Notch, since functional coupling of Notch and Serrate occurs in *Drosophila*. Two chick Notch  
30 homologs were obtained as described below.

**C-Notch-1 and C-Notch-2 are apparent counterparts of the rodent Notch-1 and Notch-2 genes, respectively**

We searched for Notch homologs in the chick by PCR,  
35 using cDNA prepared from two-day chick embryos and degenerate primers based on conserved regions common to the known rodent Notch homologs. In this way, we obtained fragments, each

approximately 1000 nucleotides long, of two distinct genes, which we have called C-Notch-1 and C-Notch-2. The fragments extend from the third Notch/lin12 repeat up to and including the last five or so EGF-like repeats. EGF-like repeats are present in a large number of proteins, most of which are otherwise unrelated to Notch. The three Notch/lin12 repeats, however, are peculiar to the Notch family of genes and are found in all its known members. C-Notch-1 shows the highest degree of amino-acid identity with rodent Notch1 (Weinmaster et al., 1991, Development 113:199-205), and is expressed in broadly similar domains to rodent Notch1 (see below). Of the rodent Notch genes, C-Notch-2 appears most similar to Notch2 (Weinmaster et al., 1992, Development 116:931-941).

We examined the expression patterns of C-Notch-1 in early embryos by *in situ* hybridization. C-Notch-1 was expressed in the 1- to 2-day chick embryo in many well-defined domains, including the neural tube, the presomitic mesoderm, the nephrogenic mesoderm (the prospective mesonephros), the nasal placode, the otic placode/vesicle, the lens placode, the epibranchial placodes, the endothelial lining of the vascular system, in the heart, and the apical ectodermal ridges (AER) of the limb buds. These sites match the reported sites of Notch1 expression in rodents at equivalent stages (Table II). Taking the sequence data together with the expression data, we conclude that C-Notch-1 is either the chick ortholog of rodent Notch1, or a very close relative of it.

Table II

30

COMPARISON OF DOMAINS OF RODENT-NOTCH1  
AND CHICK NOTCH-1 EXPRESSION THROUGHOUT EMBRYOGENESIS

	Body Region	R-Notch1'	C-Notch1
	primitive streak	+	+
35	Hensen's node	-	-
	neural tube	+	+

	retina	+	+
	lens	+	+
	otic placode/vesicle	+	+
5	epibranchial placodes	+	+
	nasal placode	+	+
	dorsal root ganglia	+	+
	presomitic mesoderm	+	+
	somites	+	+
10	notochord	?	+
	mesonephric kidney	+	+
	metanephric kidney	+	+
	blood vessels	+	+
	heart	+	+
15	whisker follicles	+	N/A
	thymus	+	?
	toothbuds	+	N/A
	salivary gland	+	?
20	limb bud (AER)	?	+
	* from Weinmaster et al., 1991, Development 113:199-205; Franco del Amo et al., 1992, Development 115:737-744; Reaume et al., 1992, Dev. Biol. 154:377-387; Kopan and Weintraub, 1993, J. Cell. Biol. 121:631-641; Lardelli et al., 1994, Mech. of Dev. 46:123-126.		

25

**C-Serrate is a homolog of *Drosophila* Serrate, and codes for a candidate ligand for a receptor belonging to the Notch family**

In *Drosophila*, two ligands for Notch are known, encoded by the two related genes *Delta* and *Serrate*. The amino-acid sequences corresponding to these genes are homologous at their 5' ends, including a region, the DSL motif, which is necessary and sufficient for in vitro binding to Notch. To isolate a fragment of a chicken homolog of *Serrate*, we used PCR and degenerate primers designed to recognize sequences on either side of the DSL motif (see Materials and methods). A 900 base pair PCR fragment was

recovered and used to screen a library, allowing us to isolate overlapping cDNA clones. The DNA sequence of the cDNA clones revealed an almost complete single open reading frame of 3582 nucleotides, lacking only a few 5' bases.

- 5 Comparison with the amino acid sequences of *Drosophila* Delta and Serrate suggests that we are missing only the portion of the coding sequence that encodes part of the signal sequence of the chick Serrate protein.

#### Translation of the nucleotide sequence

- 10 (SEQ ID NO:5) (Fig. 3) predicts a protein of 1230 amino acids (SEQ ID NO:6) (Fig. 4). A hydropathy plot reveals a single hydrophobic region characteristic of a transmembrane domain (Kyte and Doolittle, 1982, J. Mol. Biol. 157:105-132). In addition, the protein has sixteen EGF-like repeats organized  
 15 in a tandem array in its extracellular domain. Comparison of the chick sequence with sequences of *D. melanogaster* Delta and Serrate suggests that the clones encode a chicken homolog of Serrate (Fig. 5; Fig. 6). Whereas *Drosophila* Serrate contains 14 EGF-like repeats with large insertions in repeats  
 20 4, 6 and 10, the chicken homolog has an extra two EGF-like repeats and only one small insertion of 16 amino acids in the 10th repeat. Both proteins have a second cysteine-rich region between the EGF-like repeats and the transmembrane domain; the spacing of the cysteines in this region is almost  
 25 identical in the two proteins (compare  
 CX<sub>2</sub>CXCX<sub>6</sub>CX<sub>4</sub>CX<sub>15</sub>CX<sub>3</sub>CX<sub>7</sub>CX<sub>4</sub>CX<sub>3</sub>C in *Drosophila* Serrate with  
 CX<sub>2</sub>CXCX<sub>6</sub>CX<sub>4</sub>CX<sub>9</sub>CX<sub>3</sub>CX<sub>7</sub>CX<sub>4</sub>CX<sub>3</sub>C in C-Serrate). The intracellular domain of C-Serrate bears no significant homology to the intracellular domains of either *Drosophila* Delta or Serrate.

30

C-Serrate is expressed in the central nervous system, cranial placodes, nephric mesoderm, vascular system, and limb bud mesenchyme

- 35 *In situ* hybridization was performed to examine the expression of C-Serrate in whole-mount preparations during early embryogenesis, from stage 4 to stage 21, at intervals

of roughly 12 hours. Later stages were studied by *in situ* hybridization on cryosections.

The main sites of early expression of C-Serrate, as seen in whole mounts, can be grouped under five headings:  
5 central nervous system, cranial placodes, nephric mesoderm, vascular system, and limb bud mesenchyme.

#### *Central nervous system*

The first detectable expression of C-Serrate was  
10 seen in the central nervous system at stage 6 (0 somites/24 hrs), within the posterior portion of the neural plate. By stage 10 (9-11 somites/35.5 hrs), a strong stripe of expression was seen in the prospective diencephalon. Additional faint staining was seen in the hindbrain and in  
15 the prospective spinal cord.

At stage 13, there were several patches of expression in the neural tube. In the diencephalon, there was a strong triangular stripe of expression that appeared to correspond to neuromere D2. There were two patches (one on  
20 either side of the midline) on the floor of the anterior mesencephalon as well as diffuse staining in the dorsal mesencephalon. In the hindbrain and rostral spinal cord, there were two longitudinal stripes of expression on either side of the midline: one along the dorsal edge of the neural  
25 tube and a second more ventral one, adjacent to the floor plate. Both were located within the domain of (rat) *Notch 1* expression. The anterior limit of the ventral stripe was at the midbrain/hindbrain boundary. The dorsal stripe was continuous with the expression in the dorsal mesencephalon.  
30 In the anterior spinal cord, expression was more spotty, the stripes being replaced by isolated scattered cells expressing C-Serrate.

At stage 17 (58 hrs), expression in the diencephalon and midbrain was unchanged. In the hindbrain  
35 and spinal cord, there were an additional two longitudinal stripes: one midway along the dorsoventral axis and a second wider more ventral stripe; the anterior limits of these

stripes coincided with the anterior border of rhombomere 2. All four longitudinal stripes in the hindbrain continued into the spinal cord of the embryo; decreasing towards its posterior end. These stripes of expression were maintained  
5 at least up to and including stage 31 (E7). By stage 21 (84 hrs), additional expression was seen in the cerebral hemispheres and strong expression in a salt and pepper distribution of cells in the optic tectum.

#### 10 Cranial placodes

It is striking that C-Serrate is expressed in all the cranial placodes - the lens placode, the nasal placode, the otic placode/vesicle and the epibranchial placodes, as well as a patch of cranial ectoderm anterior to the otic  
15 placode that may correspond to the trigeminal placode (which is not well-defined morphologically).

In the lens placode, expression was already seen at stage 11, rapidly became very strong, and persisted at least to stage 21. Expression was weaker in the nasal placode and  
20 was only detected from stage 13. Again, expression was maintained at least until stage 21.

Likewise for the otic placode, expression began to be visible at stage 10 and was strong by early stage 11 (12-14 somites, 42.5 hours). Curiously, there was a "hole" in  
25 the otic expression domain - an anteroventral region of the placode in which the gene was not expressed. Subsequently, as the placode invaginates to form an otic vesicle, the strongest expression was seen at the anterolateral and posteromedial poles. Later still, as the otic vesicle  
30 becomes transformed into the membranous labyrinth of the inner ear, C-Serrate expression became restricted to the sensory patches.

The epibranchial expression was seen at stage 13/14 as strong staining in the ectoderm around the dorsal margins  
35 of the first and second branchial clefts. It was accompanied by expression of the gene in the deep part of the lining of



the clefts and in the endodermal lining of the branchial pouches, where the two epithelia abut one another.

Lastly, a large and strong but transient patch of expression was seen in the cranial ectoderm just anterior and ventral to the ear rudiment at stage 11. From its location, we suspect this to be, or to include, the region of the trigeminal placode.

#### *Nephric mesoderm*

- 10 Expression was detectable in the cells of the intermediate mesoderm from stage 10 and in older embryos (stage 17 to 21) in the developing mesonephric tubules.

#### *Limb buds*

- 15 C-Serrate mRNA was localized to a patch of mesenchyme at the distal end of the developing limb bud. This may suggest a role in limb growth.

#### *Other sites*

- 20 Expression was also seen in the tail bud, allantoic stalk, and possibly other tissues at late stages.

#### **All major sites of C-Serrate expression lie within domains of C-Notch-1 expression**

- 25 The conservation of the DSL domain and adjacent N-terminal region in C-Serrate suggests that it functions as a ligand for a receptor belonging to the Notch family. We thus expected to find sites where C-Serrate expression is accompanied by expression of a *Notch* gene. At such sites, overlapping or contiguous expression of the two genes can be taken as an indication that cells are communicating by Serrate-Notch signalling. We have compared the expression pattern of C-Serrate, as shown by *in situ* hybridization, with that of C-Notch-1, to discover what overlaps in fact occur, over a range of stages up to 8 days of incubation (E8). All the observed sites of C-Serrate expression indeed lay within,

or very closely adjacent to, domains of expression of C-Notch-1 (Table III).

Table III

5

COMPARISON OF C-NOTCH-1 AND  
C-SERRATE EXPRESSION AT STAGE 17a

	Body region	C-Notch-1	C-Serrate
		++ (almost everywhere)	++ (specific regions)
	brain and spinal cord	++	-
10	retina	++	++
	lens	+	++
	otic placode/vesicle	++	++
	epibranchial placodes	++	++
	nasal placode	++	-
	dorsal root ganglia	+	-
15	branchial mesenchyme	-	++ (furrows)
	branchial ectoderm	+	++ (tips of pouches)
	branchial endoderm	+	-
	presomitic mesoderm	++	-
	somites	++	-
20	notochord	++	++
	mesonephric kidney	++	++
	metanephric kidney	++	++
	blood vessels	++	++
	heart	+	++
25	limb bud (stage 21)	++ (AER)	++ (distal mesenchyme)

a Hamburger and Hamilton, 1951, J. Exp. Zool. 88:49-92.

30 Because of the importance of Notch and its partners in insect neurogenesis, it was of particular interest to us to see whether the homologous genes are involved in the development of the vertebrate CNS. C-Serrate is expressed in the CNS, and its pattern of expression shows a remarkable relationship to that of the Notch homologs.

35 We analyzed transverse sections through the spinal cord of a six day chicken embryo hybridized with C-Notch-1

and C-Serrate antisense RNA probes. C-Notch-1 was expressed throughout the luminal region as described previously; within this region, there were two small patches in which Serrate was strongly expressed.

5

### Discussion

In *Drosophila* development, cell-cell signalling via the product of the Notch gene plays a cardinal role in the final cell-fate decisions that specify the detailed pattern of differentiated cell types. This signalling pathway, in which the Notch protein has been identified as a transmembrane receptor, is best known for its role in neurogenesis: loss-of-function mutations in Notch or any of a set of other genes required for signal transmission via Notch alter cell fates in the neuroectoderm, causing cells that should have remained epidermal to become neural instead. Notch-dependent signalling is, however, as important in non-neural as in neural tissues. It regulates choices of mode of differentiation in oogenesis, in myogenesis, in formation of the Malpighian tubules and in the gut, for example, as well as in development of the retina, the peripheral sensilla, and the central nervous system. In most of these cases the signal delivered via Notch appears to mediate lateral inhibition, a type of interaction by which a cell that becomes committed to differentiate in a particular way - for example, as a neuroblast - inhibits its immediate neighbors from doing likewise. This forces adjacent cells to behave in contrasting ways, creating a fine-grained pattern of different cell types.

30 There are, however, good reasons to believe that this is not the only function of signals delivered via Notch. Two direct ligands of Notch have been identified. These are the products of the Delta and Serrate genes. Both of them, like Notch itself, code for transmembrane proteins with tandem arrays of EGF-like repeats in their extracellular domain. Both the Delta and the Serrate protein have been shown to bind to Notch in a cell adhesion assay, and they

share a large region of homology at their amino-termini including a motif that is necessary and sufficient for interaction with Notch in vitro, the so-called EBD or DSL domain. Yet despite these biochemical similarities, they  
5 seem to have quite different developmental functions.

Although Serrate is expressed in many sites in the fly, it is apparently required only in the humeral, wing and halteres disks. When Serrate function is lost by mutation, these structures fail to grow. Studies on the wing disc have  
10 indicated that it is specifically the wing margin that depends on Serrate; when Serrate is lacking, this critical signaling region and growth centre fails to form, and when Serrate is expressed ectopically under a GAL4-UAS promoter in the ventral part of the wing disc, ectopic wing margin tissue  
15 is induced, leading to ectopic outgrowths. Notch appears to be the receptor for Serrate at the wing margin, since some mutant alleles of Notch cause similar disturbances of wing margin development and allele-specific interactions are seen in the effects of the two genes.

20 Here we describe the identification and full length sequence of a homolog of the *Drosophila* gene Serrate, and identification and partial sequence of chick homologs of rat/mouse *Notch1* and *Notch2*.

Within the chick Serrate cDNA there is a single  
25 open reading frame predicted to encode a large transmembrane protein with 16 EGF repeats in its extracellular domain. It has a well conserved DSL motif suggesting that it would interact directly with Notch. The intracellular domain of chick Serrate exhibits no homology to anything in the current  
30 databases including the intracellular domains of *Drosophila* Delta and Serrate. It should be pointed out however that the intracellular domains of chick and human Serrate (see Section 12) are almost identical.

The spatial distributions of C-Notch-1 and  
35 C-Serrate were investigated during early embryogenesis by in situ hybridization. C-Notch-1 and C-Serrate exhibit dynamic and complex patterns of expression including several regions

in which they are coexpressed (CNS, ear, branchial region, lens, heart, nasal placodes and mesonephros). The overlapping expression together with the finding that C-Serrate has a well conserved Notch binding domain suggests  
5 that this receptor/ligand interaction has been conserved from *Drosophila* through to vertebrates.

In *Drosophila*, the Notch receptor is quite widely distributed and its ligands are found in overlapping but more restricted domains. In the chick a similar situation is  
10 observed.

Fly Notch is necessary for many steps in the development of *Drosophila*; its role in lateral inhibition especially in the development of the central nervous system and peripheral sense organs being the best studied examples.  
15 However, Notch is a multifunctional receptor and can interact with different signalling molecules (including Delta and Serrate) and in developmental processes that do not easily fit within the framework of lateral inhibition. While available evidence implicates Delta as the signalling  
20 molecule in lateral inhibition there is no data to suggest that Serrate participates in lateral inhibition. Rather, Serrate appears to be necessary for development of the dorsal imaginal discs of the larva; that is, the humeral, haltere and wing discs. In the latter, the best studied of these  
25 processes, Serrate and Notch are important for the development of the dorsoventral wing margin, a structure necessary for the organization of wing development as a whole.

That C-Serrate has a significant function can be  
30 inferred from the conservation of its sequence, in particular, of its Notch-binding domain. The expression patterns reported for C-Serrate in this paper provide the following information. First, since the Serrate gene is expressed in or next to sites where C-Notch-1 is expressed  
35 (possibly in conjunction with other Notch homologs), it is highly probable that C-Serrate exerts its action by binding to C-Notch-1 (or to another chick Notch homolog with a

similar expression pattern). Second, the expression in the developing kidney, the vascular system and the limb buds might reflect an involvement in inductive signalling between mesoderm and ectoderm, which plays an important part in the development of all these organs. In the limb buds, for example, C-Serrate is expressed in the distal mesoderm, and C-Notch-1 is expressed in the overlying apical ectodermal ridge, whose maintenance is known to depend on a signal from the mesoderm below. In the cranial placodes, a similar role is possible, but the evidence for inductive signalling is weaker, and C-Serrate may equally be involved in communications between cells within the placodal epithelium, for example, in regulating the specialized modes of differentiation of the placodal cells.

What might C-Serrate's function be within the curiously restricted domains of its expression in the CNS? One possibility is that it is involved in regulating the production of oligodendrocytes, which have likewise been reported to originate from narrow bands of tissue extending along the cranio-caudal axis of the neural tube.

#### 9. ISOLATION AND CHARACTERIZATION OF HUMAN SERRATE HOMOLOGS

Clones for the human Serrate sequence were obtained as described below.

The polymerase chain reaction (PCR) was used to amplify DNA from a human placenta cDNA library. Degenerate oligonucleotide primers used in this reaction were designed based on amino-terminal regions of high homology between *Drosophila* Serrate and *Drosophila* Delta (see Fig. 5); this high homology region includes the 5' "DSL" domain, that is believed to code for the Notch-binding portion of Delta and Serrate. Two PCR products were isolated and used, one a 350 bp fragment, and one a 1.2 kb fragment. These PCR fragments were labeled with <sup>32</sup>P and used to screen a commercial human fetal brain cDNA library made from a 17-18 week old fetus

(previously available from Stratagene), in which the cDNAs were inserted into the *EcoRI* site of a  $\lambda$ -Zap vector.

The 1.2 kb fragment hybridized to a single clone out of the  $10^6$  clones screened. We rescued this fragment from the  $\lambda$  DNA by converting the isolated phage  $\lambda$  clone to a plasmid via the manufacturer's instructions, yielding the Serrate-homologous cDNA as an insert in the *EcoRI* site of the vector Bluescript KS- (Stratagene). This plasmid was named "pBS39" and the gene corresponding to this cDNA clone was called *Human Serrate-1* (also known as *Human Jagged-1* ("HJ1")). The isolated cDNA was 6464 nucleotides long and contained a complete open reading frame as well as 5' and 3' untranslated regions (Fig. 1). Sequencing was carried out using the Sequenase® sequencing system (U.S. Biochemical Corp.) on 5 and 6% Sequagel acrylamide sequencing gels.

The 350 bp fragment hybridized with two clones, containing cDNA inserts of approximately 1.1 and 3.1 kb in length; the plasmid constructs containing these inserts were named pBS14 and pBS15, respectively. Each clone was isolated, its respective insert rescued from the  $\lambda$  cDNA, and sequenced as above. The nucleotide sequence of the pBS14 insert was identical to a 1.1 kb stretch of sequence contained internally within the pBS15 cDNA insert and therefore, this clone was not characterized further. The sequence of the 3.1 kb pBS15 insert encoded a single open reading frame which spanned all but the 5' 20 nucleotides of the insert. The methionine located at the amino terminal residue of this predicted open reading was homologous to the start methionine encoded by the *Human Serrate-1* (HJ1) cDNA clone in pBS39. The gene encoding the cDNA insert of pBS15 was named *Human Serrate-2* and is also known as *Human Jagged-2* ("HJ2").

The pBS15 (HJ2) 3.1 kb insert was then labeled with  $^{32}\text{P}$  and used to screen another human fetal brain library (from Clontech), in which cDNA generated from a 25-26 week-old fetus was cloned into the *EcoRI* site of  $\lambda$ gt11. This screen identified three potential positive clones. To isolate the

cdNAs,  $\lambda$ gt11 DNA was prepared from a liquid lysate and purified over a DEAE column. The purified DNA was then cut with *EcoRI* and the cDNA inserts were isolated and subcloned into the *EcoRI* site of Bluescript KS-. The bluescript  
5 constructs containing these cdNAs were named pBS3-15, pBS3-2, and pBS3-20. Two of these cDNA clones, pBS3-2 and pBS3-20, contained sequences that partially overlapped with pBS15 and were further characterized. pBS3-2 had a 3.2 kb insert extending from nucleotide 1210 of the pBS15 cDNA insert to  
10 just after the polyadenylation signal. The 2.6 kb insert of pBS3-20, was restriction mapped and partially sequenced to determine its 3' and 5' ends. This analysis indicated that the pBS3-20 insert had a nucleic acid sequence that was fully contained within the pBS3-2 cDNA insert and therefore, the  
15 pBS3-20 insert was not characterized further. The insert of pBS3-15 was determined to be a Bluescript vector fragment contaminant.

Alignment of the deduced amino acid sequence (SEQ ID NO:4) of the "complete" Human Serrate-2 (HJ2) cDNA  
20 (SEQ ID NO:3) generated on the computer with the deduced amino acid sequence of Human Serrate-1 (HJ1) from pBS39 (SEQ ID NO:2) revealed a gap of about 120 bases, leading to a frameshift, in the region encoded by the pBS15 (HJ2) insert, between the putative signal sequence and the beginning of the  
25 DSL domain (Fig. 2). The nucleotides missing in the gap of the pBS15 insert would be located between nucleotides 240 and 241 of SEQ ID NO:3. This missing region probably resulted from a cloning artifact in the construction of the Stratagene library.

30 Attempts to clone the 5' end of HJ2 using anchored PCR, RACE, and Takara extended PCR techniques were unsuccessful. However, three human genomic clones potentially containing the 5' end of HJ2 were obtained from the screening of a human genomic cosmid library in which  
35 kb fragments were cloned into a unique *XhoI* site introduced into the *BamHI* site of a pWE15 vector (the unmodified vector is available from Stratagene). This cosmid library was



screened with a PCR fragment that had been amplified from the 5' end of pBS15 (HJ2) and three positive cosmid clones were isolated. Two different sets of primers were used to amplify DNA corresponding to the 5' end of pBS15 using the cosmid clones as a template, and both sets generated single bands that were subcloned, but which were determined to contain PCR artifacts. Portions of the cosmid clones are being subcloned directly without PCR, in order to obtain a portion of the cosmid clones that contains the 120 nucleotide stretch of DNA that is missing from pBS15.

The pBS39 cDNA insert, encoding the *Human Serrate-1* homolog (HJ1), has been sequenced and contains the complete coding sequence for the gene product. The nucleotide (SEQ ID NO:1) and protein (SEQ ID NO:2) sequences are shown in Figure 1. The nucleotide sequence of *Human Serrate-1* (HJ1) was translated using MacVector software (International Biotechnology Inc., New Haven, CT). The coding region consists of nucleotide numbers 371-4024 of SEQ ID NO:1. The Protean protein analysis software program from DNASTar (Madison, WI) was used to predict signal peptide and transmembrane regions (based on hydrophobicity). The signal peptide was predicted to consist of amino acids 14-29 of SEQ ID NO:2 (encoded by nucleotide numbers 410-457 of SEQ ID NO:1), whereby the amino terminus of the mature protein was predicted to start with Gly at amino acid number 30. The transmembrane domain was predicted to be amino acid numbers 1068-1089 of SEQ ID NO:2, encoded by nucleotide numbers 3572-3637 of SEQ ID NO:1. The consensus (DSL) domain, the region of homology with *Drosophila* Delta and Serrate, predicted to mediate binding with Notch (in particular, Notch ELR 11 and 12), spans amino acids 185-229 of SEQ ID NO:2, encoded by nucleotide numbers 923-1057 of SEQ ID NO:1. Epidermal growth factor-like (ELR) repeats in the amino acid sequence were identified by eye; 15 (full-length) ELRs were identified and 3 partial ELRs as follows:

ELR 1: amino acid numbers 234 - 264

ELR 2: amino acid numbers 265 - 299

- ELR 3: amino acid numbers 300 - 339  
 ELR 4: amino acid numbers 340 - 377  
 ELR 5: amino acid numbers 378 - 415  
 ELR 6: amino acid numbers 416 - 453  
 5 ELR 7: amino acid numbers 454 - 490  
 ELR 8: amino acid numbers 491 - 528  
 ELR 9: amino acid numbers 529 - 566  
 Partial ELR: amino acid numbers 567 - 598  
 Partial ELR: amino acid numbers 599 - 632  
 10 ELR 10: amino acid numbers 633 - 670  
 ELR 11: amino acid numbers 671 - 708  
 ELR 12: amino acid numbers 709 - 747  
 ELR 13: amino acid numbers 748 - 785  
 ELR 14: amino acid numbers 786 - 823  
 15 ELR 15: amino acid numbers 824 - 862  
 Partial ELR: amino acid numbers 863 - 879  
 Partial ELR: amino acid numbers 880 - 896

The total ELR domain is thus amino acid numbers 234 - 896  
 (encoded by nucleotide numbers 1070 - 3058 of SEQ ID NO:1).

- 20 The extracellular domain is thus predicted to be amino acid  
 numbers 1 - 1067 of SEQ ID NO:2, encoded by nucleotide  
 numbers 371 - 3571 of SEQ ID NO:1 (amino acid numbers  
 30 - 1067 in the mature protein; encoded by nucleotides  
 number 458 - 3571 of SEQ ID NO:1). The intracellular  
 25 (cytoplasmic) domain is thus predicted to be amino acid  
 numbers 1090 - 1218 of SEQ ID NO:2, encoded by nucleotide  
 numbers 3638 - 4024 of SEQ ID NO:1.

- The expression of *HJ1* in certain human tissues was  
 established by probing a Clontech Human Multiple Tissue  
 30 Northern blot with radio-labeled pBS39. The probe hybridized  
 to a single band of about 6.6 kb, and was expressed in all of  
 the tissue assayed, which included, heart, brain, placenta,  
 lung, skeletal muscle, pancreas, liver and kidney. The  
 observation that *HJ1* was expressed in adult skeletal and  
 35 heart muscle was particularly interesting, because adult  
 muscle fibers are completely surrounded by a lamina of  
 extracellular matrix, and it is unlikely, therefore, that the

role of HJ1 in these cells is in direct cell-cell communication.

- The "complete" (containing an internal deletion) *Human Serrate-2 (HJ2)* cDNA nucleotide sequence (SEQ ID NO:3) and amino acid sequence (SEQ ID NO:4) generated on the computer are shown in Figure 2. The nucleotide sequence translated using MacVector software (International Biotechnology Inc., New Haven, CT). The coding region consists of nucleotides number 332 - 4102 of SEQ ID NO:3.
- 10 The Protean protein analysis software program from DNASTar (Madison, WI) was used to predict signal peptide and transmembrane regions (based on hydrophobicity). The transmembrane domain was predicted to be amino acid numbers 912-933 of SEQ ID NO:4, encoded by nucleotides numbers 3065-3130 of SEQ ID NO:3. The consensus (DSL) domain, the region of homology with *Drosophila* Delta and Serrate, predicted to mediate binding with Notch (in particular, Notch ELR 11 and 12), spans amino acids 26-70 of SEQ ID NO:4, encoded by nucleotide numbers 407 - 541 of SEQ ID NO:3.
- 20 Epidermal growth factor-like (ELR) repeats in the amino acid sequence were identified by eye; 15 (full-length) ELRs were identified and 3 partial ELRs as follows:
- ELR 1: amino acid numbers 75 - 105
  - ELR 2: amino acid numbers 106 - 140
  - 25 ELR 3: amino acid numbers 141 - 180
  - ELR 4: amino acid numbers 181 - 218
  - ELR 5: amino acid numbers 219 - 256
  - ELR 6: amino acid numbers 257 - 294
  - ELR 7: amino acid numbers 295 - 331
  - 30 ELR 8: amino acid numbers 332 - 369
  - ELR 9: amino acid numbers 370 - 407
  - Partial ELR: amino acid numbers 408 - 435
  - Partial ELR: amino acid numbers 436 - 469
  - ELR 10: amino acid numbers 470 - 507
  - 35 ELR 11: amino acid numbers 508 - 545
  - ELR 12: amino acid numbers 546 - 584
  - ELR 13: amino acid numbers 585 - 622

ELR 14: amino acid numbers 623 - 660

ELR 15: amino acid numbers 664 - 701

Partial ELR: amino acid numbers 702 - 718

Partial ELR: amino acid numbers 719 - 735

5 The total ELR domain is thus amino acid numbers 75 - 735  
(encoded by nucleotides number 554 - 2536 of SEQ ID NO:3).  
The extracellular domain is thus predicted to be amino acid  
numbers 1 - 912 of SEQ ID NO:4, encoded by nucleotides number  
332 - 3064 of SEQ ID NO:3. The intracellular (cytoplasmic)  
10 domain is thus predicted to be amino acid numbers 934 - 1257  
of SEQ ID NO:4, encoded by nucleotide numbers 3131 - 4102 of  
SEQ ID NO:3.

Like *Human Serrate-1 (HJ1)*, the "complete" (with an  
internal deletion) *Human Serrate-2 (HJ2)* cDNA (SEQ ID NO:3)  
15 generated on the computer encodes a protein containing 16  
complete and 2 interrupted EGF repeats as well as the  
diagnostic cryptic EGF repeat known as the DSL domain, which  
has been found only in putative Notch ligands. The open  
reading frame of the computer generated "complete" Human  
20 *Serrate-2 (HJ2)* is about 1400 amino acids long, approximately  
182 amino acids longer than the carboxy terminus of HJ1 and  
the rat *Serrate* homologue *Jagged*. While there is significant  
homology between the complete HJ2 and HJ1 in the amino  
terminal portion of the protein, this homology is lost just  
25 before the putative transmembrane domain at about amino acid  
number 1029 of HJ1. This result is particularly interesting  
because the presence of a long COOH-terminal tail implies the  
possibility of some additional function or regulation of HJ2.

The "complete" (with an internal deletion) *Human*  
30 *Serrate-2 (HJ2)* cDNA (SEQ ID NO:3) sequence can be  
constructed by taking advantage of the unique restriction  
sites for *AccI*, *DraIII*, or *BamHI* present in the sequence  
overlap of pBS15 and pBS3-2, and which enzymes cleave the  
pBS15 insert at nucleotides 1431, 2648, and 2802,  
35 respectively.

The expression of *HJ2* in certain human tissues was established by probing a Clontech Human Multiple Tissue Northern blot with radio-labeled clone pBS15. This probe hybridized to a single band of about 5.2 kb and was expressed  
5 in heart, brain, placenta, lung, skeletal muscle, and pancreas, but was absent or nearly undetectable in liver and kidney. As in the case of *HJ1* expression discussed supra, the observation that the pBS15 insert component of *HJ2* was expressed in adult skeletal and heart muscle was particularly  
10 interesting, because adult muscle fibers are completely surrounded by a lamina of extracellular matrix, and it is unlikely, therefore, that the role of *HJ2* in these cells is in direct cell-cell communication.

Expression constructs are made using the isolated  
15 clone(s). The clone is excised from its vector as an *EcoRI* restriction fragment(s) and subcloned into the *EcoRI* restriction site of an expression vector. This allows for the expression of the Human Serrate protein product from the subclone in the correct reading frame. Using this  
20 methodology, expression constructs in which the *HJ1* cDNA insert of pBS39 was cloned into an expression vector for expression under the control of a cytomegalovirus promoter have been generated and *HJ1* has been expressed in both 3T3 and HAKAT human keratinocyte cell lines.

25

#### 10. DEPOSIT OF MICROORGANISMS

Plasmid pBS39, containing an *EcoRI* fragment encoding full-length Human Serrate-1 (*HJ1*), was deposited on February 28, 1995 with the American Type Culture Collection,  
30 1201 Parklawn Drive, Rockville, Maryland 20852, under the provisions of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedures, and assigned Accession No. 97068.

Plasmid pBS15, containing a 3.1 kb *EcoRI* fragment  
35 encoding the amino terminus of Human Serrate-2 (*HJ2*), cloned into the *EcoRI* site of Bluescript KS-, was deposited on March 5, 1996 with the American Type Culture Collection, 1201

Parklawn Drive, Rockville, Maryland 20852, under the provisions of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedures, and assigned Accession No. \_\_\_\_\_.

5           Plasmid pBS3-2 containing an 3.2 kb *EcoRI* fragment encoding the carboxy terminus of Human Serrate-2 (HJ2), cloned into the *EcoRI* site of Bluescript KS-, was deposited on March 5, 1996 with the American Type Culture Collection, 1201 Parklawn Drive, Rockville, Maryland 20852, under the  
10 provisions of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedures, and assigned Accession No. \_\_\_\_\_.

The present invention is not to be limited in scope  
15 by the microorganisms deposited or the specific embodiments described herein. Indeed, various modifications of the invention in addition to those described herein will become apparent to those skilled in the art from the foregoing description and accompanying figures. Such modifications are  
20 intended to fall within the scope of the appended claims.

Various references are cited herein, the disclosures of which are incorporated by reference in their entireties.

25

30

35

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

(i) APPLICANT: ISH-HOROWICZ, DAVID  
HENRIQUE, DOMINGOS MANUEL PINTO  
LEWIS, JULIAN HART  
MYAT, ANNA MARY  
ARTAVANIS-TSAKONAS, SPYRIDON  
MANN, ROBERT S.  
GRAY, GRACE E.

(ii) TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF VERTEBRATE  
SERRATE GENES AND METHODS BASED THEREON

(iii) NUMBER OF SEQUENCES: 18

## (iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Pennie & Edmonds  
(B) STREET: 1155 Avenue of the Americas  
(C) CITY: New York  
(D) STATE: New York  
(E) COUNTRY: USA  
(F) ZIP: 10036-2711

## (v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk  
(B) COMPUTER: IBM PC compatible  
(C) OPERATING SYSTEM: PC-DOS/MS-DOS  
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

## (vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: To Be Assigned  
(B) FILING DATE: On Even Date Herewith  
(C) CLASSIFICATION:

## (viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Misrock, S. Leslie  
(B) REGISTRATION NUMBER: 18,872  
(C) REFERENCE/DOCKET NUMBER: 7326-037-228

## (ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (212) 790-9090  
(B) TELEFAX: (212) 869-9741/8864  
(C) TELEX: 66141 PENNIE

## (2) INFORMATION FOR SEQ ID NO:1:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6464 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

## (ix) FEATURE:

(A) NAME/KEY: CDS  
(B) LOCATION: 371..4027

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GAATTCCCCT CCCCCCTTTT TCCATGCAGC TGATCTAAAA GGGAATAAAA GGCTGCGCAT	60
AATCATAATA ATAAAAGAAG GGGAGCGCGA GAGAAGGAAA GAAAGCCGGG AGGTGGAAGA	120
GGAGGGGGAG CGTCTCAAAG AAGCGATCAG AATAATAAAA GGAGGCCGGG CTCTTTGCCT	180
TCTGGAACGG GCCGCTCTTG AAAGGGCTTT TGAAAAGTGG TGTGTTTTTC CAGTCGTGCA	240
TGCTCCAATC GCGCGAGTAT ATTAGAGCCG GGACGCGGCC GCAGGGGCAG CGGCGACGGC	300
AGCACCGGCG GCAGCACCAG CGCGAACAGC AGCGGCGGCG TCCCGAGTGC CCGCGGCGGC	360
GCGCGCAGCG ATG CGT TCC CCA CGG ACA CGC GGC CGG TCC GGG CGC CCC	409
Met Arg Ser Pro Arg Thr Arg Gly Arg Ser Gly Arg Pro	
1 5 10	
CTA AGC CTC CTG CTC GCC CTG CTC TGT GCC CTG CGA GCC AAG GTG TGT	457
Leu Ser Leu Leu Leu Ala Leu Leu Cys Ala Leu Arg Ala Lys Val Cys	
15 20 25	
GGG GCC TCG GGT CAG TTC GAG TTG GAG ATC CTG TCC ATG CAG AAC GTG	505
Gly Ala Ser Gly Gln Phe Glu Leu Glu Ile Leu Ser Met Gln Asn Val	
30 35 40 45	
AAC GGG GAG CTG CAG AAC GGG AAC TGC TGC GGC GGC GCC CGG AAC CCG	553
Asn Gly Glu Leu Gln Asn Gly Asn Cys Cys Gly Gly Ala Arg Asn Pro	
50 55 60	
GGA GAC CGC AAG TGC ACC CGC GAC GAG TGT GAC ACA TAC TTC AAA GTG	601
Gly Asp Arg Lys Cys Thr Arg Asp Glu Cys Asp Thr Tyr Phe Lys Val	
65 70 75	
TGC CTC AAG GAG TAT CAG TCC CGC GTC ACG GCC GGG GGG CCC TGC AGC	649
Cys Leu Lys Glu Tyr Gln Ser Arg Val Thr Ala Gly Gly Pro Cys Ser	
80 85 90	
TTC GGC TCA GGG TCC ACG CCT GTC ATC GGG GGC AAC ACC TTC AAC CTC	697
Phe Gly Ser Gly Ser Thr Pro Val Ile Gly Gly Asn Thr Phe Asn Leu	
95 100 105	
AAG GCC AGC CGC GGC AAC GAC CCG AAC CGC ATC GTG CTG CCT TTC AGT	745
Lys Ala Ser Arg Gly Asn Asp Pro Asn Arg Ile Val Leu Pro Phe Ser	
110 115 120 125	
TTC GCC TGG CCG AGG TCC TAT ACG TTG CTT GTG GAG GCG TGG GAT TCC	793
Phe Ala Trp Pro Arg Ser Tyr Thr Leu Val Glu Ala Trp Asp Ser	
130 135 140	
AGT AAT GAC ACC GTT CAA CCT GAC AGT ATT ATT GAA AAG GCT TCT CAC	841
Ser Asn Asp Thr Val Gln Pro Asp Ser Ile Ile Glu Lys Ala Ser His	
145 150 155	
TCG GGC ATG ATC AAC CCC AGC CGG CAG TGG CAG ACG CTG AAG CAG AAC	889
Ser Gly Met Ile Asn Pro Ser Arg Gln Trp Gln Thr Leu Lys Gln Asn	
160 165 170	
ACG GGC GTT GCC CAC TTT GAG TAT CAG ATC CGC GTG ACC TGT GAT GAC	937
Thr Gly Val Ala His Phe Glu Tyr Gln Ile Arg Val Thr Cys Asp Asp	
175 180 185	
TAC TAC TAT GGC TTT GGC TGT AAT AAG TTC TGC CGC CCC AGA GAT GAC	985
Tyr Tyr Tyr Gly Phe Gly Cys Asn Lys Phe Cys Arg Pro Arg Asp Asp	
190 195 200 205	
TTC TTT GGA CAC TAT GCC TGT GAC CAG AAT GGC AAC AAA ACT TGC ATG	1033



Phe	Phe	Gly	His	Tyr	Ala	Cys	Asp	Gln	Asn	Gly	Asn	Lys	Thr	Cys	Met	
				210					215					220		
GAA	GGC	TGG	ATG	GGC	CCC	GAA	TGT	AAC	AGA	GCT	ATT	TGC	CGA	CAA	GGC	1081
Glu	Gly	Trp	Met	Gly	Pro	Glu	Cys	Asn	Arg	Ala	Ile	Cys	Arg	Gln	Gly	
			225					230				235				
TGC	AGT	CCT	AAG	CAT	GGG	TCT	TGC	AAA	CTC	CCA	GGT	GAC	TGC	AGG	TGC	1129
Cys	Ser	Pro	Lys	His	Gly	Ser	Cys	Lys	Leu	Pro	Gly	Asp	Cys	Arg	Cys	
			240				245					250				
CAG	TAC	GGC	TGG	CAA	GGC	CTG	TAC	TGT	GAT	AAG	TGC	ATC	CCA	CAC	CCG	1177
Gln	Tyr	Gly	Trp	Gln	Gly	Leu	Tyr	Cys	Asp	Lys	Cys	Ile	Pro	His	Pro	
	255					260					265					
GGA	TGC	GTC	CAC	GGC	ATC	TGT	AAT	GAG	CCC	TGG	CAG	TGC	CTC	TGT	GAG	1225
Gly	Cys	Val	His	Gly	Ile	Cys	Asn	Glu	Pro	Trp	Gln	Cys	Leu	Cys	Glu	
	270				275					280					285	
ACC	AAC	TGG	GGC	GGC	CAG	CTC	TGT	GAC	AAA	GAT	CTC	AAT	TAC	TGT	GGG	1273
Thr	Asn	Trp	Gly	Gly	Gln	Leu	Cys	Asp	Lys	Asp	Leu	Asn	Tyr	Cys	Gly	
				290					295					300		
ACT	CAT	CAG	CCG	TGT	CTC	AAC	GGG	GGA	ACT	TGT	AGC	AAC	ACA	GGC	CCT	1321
Thr	His	Gln	Pro	Cys	Leu	Asn	Gly	Gly	Thr	Cys	Ser	Asn	Thr	Gly	Pro	
			305					310					315			
GAC	AAA	TAT	CAG	TGT	TCC	TGC	CCT	GAG	GGG	TAT	TCA	GGA	CCC	AAC	TGT	1369
Asp	Lys	Tyr	Gln	Cys	Ser	Cys	Pro	Glu	Gly	Tyr	Ser	Gly	Pro	Asn	Cys	
		320					325					330				
GAA	ATT	GCT	GAG	CAC	GCC	TGC	CTC	TCT	GAT	CCC	TGT	CAC	AAC	AGA	GGC	1417
Glu	Ile	Ala	Glu	His	Ala	Cys	Leu	Ser	Asp	Pro	Cys	His	Asn	Arg	Gly	
	335					340					345					
AGC	TGT	AAG	GAG	ACC	TCC	CTG	GGC	TTT	GAG	TGT	GAG	TGT	TCC	CCA	GGC	1465
Ser	Cys	Lys	Glu	Thr	Ser	Leu	Gly	Phe	Glu	Cys	Glu	Cys	Ser	Pro	Gly	
	350				355					360					365	
TGG	ACC	GGC	CCC	ACA	TGC	TCT	ACA	AAC	ATT	GAT	GAC	TGT	TCT	CCT	AAT	1513
Trp	Thr	Gly	Pro	Thr	Cys	Ser	Thr	Asn	Ile	Asp	Asp	Cys	Ser	Pro	Asn	
				370					375					380		
AAC	TGT	TCC	CAC	GGG	GGC	ACC	TGC	CAG	GAC	CTG	GTT	AAC	GGA	TTT	AAG	1561
Asn	Cys	Ser	His	Gly	Gly	Thr	Cys	Gln	Asp	Leu	Val	Asn	Gly	Phe	Lys	
			385					390					395			
TGT	GTG	TGC	CCC	CCA	CAG	TGG	ACT	GGG	AAA	ACG	TGC	CAG	TTA	GAT	GCA	1609
Cys	Val	Cys	Pro	Pro	Gln	Trp	Thr	Gly	Lys	Thr	Cys	Gln	Leu	Asp	Ala	
		400					405					410				
AAT	GAA	TGT	GAG	GCC	AAA	CCT	TGT	GTA	AAC	GCC	AAA	TCC	TGT	AAG	AAT	1657
Asn	Glu	Cys	Glu	Ala	Lys	Pro	Cys	Val	Asn	Ala	Lys	Ser	Cys	Lys	Asn	
	415					420					425					
CTC	ATT	GCC	AGC	TAC	TAC	TGC	GAC	TGT	CTT	CCC	GGC	TGG	ATG	GGT	CAG	1705
Leu	Ile	Ala	Ser	Tyr	Tyr	Cys	Asp	Cys	Leu	Pro	Gly	Trp	Met	Gly	Gln	
	430				435					440					445	
AAT	TGT	GAC	ATA	AAT	ATT	AAT	GAC	TGC	CTT	GGC	CAG	TGT	CAG	AAT	GAC	1753
Asn	Cys	Asp	Ile	Asn	Ile	Asn	Asp	Cys	Leu	Gly	Gln	Cys	Gln	Asn	Asp	
				450				455						460		
GCC	TCC	TGT	CGG	GAT	TTG	GTT	AAT	GGT	TAT	CGC	TGT	ATC	TGT	CCA	CCT	1801
Ala	Ser	Cys	Arg	Asp	Leu	Val	Asn	Gly	Tyr	Arg	Cys	Ile	Cys	Pro	Pro	
			465					470					475			

GGC TAT GCA GGC GAT CAC TGT GAG AGA GAC ATC GAT GAA TGT GCC ACC Gly Tyr Ala Gly Asp His Cys Glu Arg Asp Ile Asp Glu Cys Ala Ser 480 485 490	1849
AAC CCC TGT TTG AAT GGG GGT CAC TGT CAG AAT GAA ATC AAC AGA TTC Asn Pro Cys Leu Asn Gly Gly His Cys Gln Asn Glu Ile Asn Arg Phe 495 500 505	1897
CAG TGT CTG TGT CCC ACT GGT TTC TCT GGA AAC CTC TGT CAG CTG GAC Gln Cys Leu Cys Pro Thr Gly Phe Ser Gly Asn Leu Cys Gln Leu Asp 510 515 520 525	1945
ATC GAT TAT TGT GAG CCT AAT CCC TGC CAG AAC GGT GCC CAG TGC TAC Ile Asp Tyr Cys Glu Pro Asn Pro Cys Gln Asn Gly Ala Gln Cys Tyr 530 535 540	1993
AAC CGT GCC AGT GAC TAT TTC TGC AAG TGC CCC GAG GAC TAT GAG GGC Asn Arg Ala Ser Asp Tyr Phe Cys Lys Cys Pro Glu Asp Tyr Glu Gly 545 550 555	2041
AAG AAC TGC TCA CAC CTG AAA GAC CAC TGC CGC ACG ACC CCC TGT GAA Lys Asn Cys Ser His Leu Lys Asp His Cys Arg Thr Thr Pro Cys Glu 560 565 570	2089
GTG ATT GAC AGC TGC ACA GTG GCC ATG GCT TCC AAC GAC ACA CCT GAA Val Ile Asp Ser Cys Thr Val Ala Met Ala Ser Asn Asp Thr Pro Glu 575 580 585	2137
GGG GTG CGG TAT ATT TCC TCC AAC GTC TGT GGT CCT CAC GGG AAG TGC Gly Val Arg Tyr Ile Ser Ser Asn Val Cys Gly Pro His Gly Lys Cys 590 595 600 605	2185
AAG AGT CAG TCG GGA GGC AAA TTC ACC TGT GAC TGT AAC AAA GGC TTC Lys Ser Gln Ser Gly Gly Lys Phe Thr Cys Asp Cys Asn Lys Gly Phe 610 615 620	2233
ACG GGA ACA TAC TGC CAT GAA AAT ATT AAT GAC TGT GAG AGC AAC CCT Thr Gly Thr Tyr Cys His Glu Asn Ile Asn Asp Cys Glu Ser Asn Pro 625 630 635	2281
TGT AGA AAC GGT GGC ACT TGC ATC GAT GGT GTC AAC TCC TAC AAG TGC Cys Arg Asn Gly Gly Thr Cys Ile Asp Gly Val Asn Ser Tyr Lys Cys 640 645 650	2329
ATC TGT AGT GAC GGC TGG GAG GGG GCC TAC TGT GAA ACC AAT ATT AAT Ile Cys Ser Asp Gly Trp Glu Gly Ala Tyr Cys Glu Thr Asn Ile Asn 655 660 665	2377
GAC TGC AGC CAG AAC CCC TGC CAC AAT GGG GGC ACG TGT CGC GAC CTG Asp Cys Ser Gln Asn Pro Cys His Asn Gly Gly Thr Cys Arg Asp Leu 670 675 680 685	2425
GTC AAT GAC TTC TAC TGT GAC TGT AAA AAT GGG TGG AAA GGA AAG ACC Val Asn Asp Phe Tyr Cys Asp Cys Lys Asn Gly Trp Lys Gly Lys Thr 690 695 700	2473
TGC CAC TCA CGT GAC AGT CAG TGT GAT GAG GCC ACG TGC AAC AAC GGT Cys His Ser Arg Asp Ser Gln Cys Asp Glu Ala Thr Cys Asn Asn Gly 705 710 715	2521
GGC ACC TGC TAT GAT GAG GGG GAT GCT TTT AAG TGC ATG TGT CCT GGC Gly Thr Cys Tyr Asp Glu Gly Asp Ala Phe Lys Cys Met Cys Pro Gly 720 725 730	2569
GGC TGG GAA GGA ACA ACC TGT AAC ATA GCC CGA AAC AGT AGC TGC CTG Gly Trp Glu Gly Thr Thr Cys Asn Ile Ala Arg Asn Ser Ser Cys Leu 735 740 745	2617

CCC AAC CCC TGC CAT AAT GGG GGC ACA TGT GTG GTC AAC GGC GAG TCC Pro Asn Pro Cys His Asn Gly Gly Thr Cys Val Val Asn Gly Glu Ser 750 755 760 765	2665
TTT ACG TGC GTC TGC AAG GAA GGC TGG GAG GGG CCC ATC TGT GCT CAG Phe Thr Cys Val Cys Lys Glu Gly Trp Glu Gly Pro Ile Cys Ala Gln 770 775 780	2713
AAT ACC AAT GAC TGC AGC CCT CAT CCC TGT TAC AAC AGC GGC ACC TGT Asn Thr Asn Asp Cys Ser Pro His Pro Cys Tyr Asn Ser Gly Thr Cys 785 790 795	2761
GTG GAT GGA GAC AAC TGG TAC CGG TGC GAA TGT GCC CCG GGT TTT GCT Val Asp Gly Asp Asn Trp Tyr Arg Cys Glu Cys Ala Pro Gly Phe Ala 800 805 810	2809
GGG CCC GAC TGC AGA ATA AAC ATC AAT GAA TGC CAG TCT TCA CCT TGT Gly Pro Asp Cys Arg Ile Asn Ile Asn Glu Cys Gln Ser Ser Pro Cys 815 820 825	2857
GCC TTT GGA GCG ACC TGT GTG GAT GAG ATC AAT GGC TAC CGG TGT GTC Ala Phe Gly Ala Thr Cys Val Asp Glu Ile Asn Gly Tyr Arg Cys Val 830 835 840 845	2905
TGC CCT CCA GGG CAC AGT GGT GCC AAG TGC CAG GAA GTT TCA GGG AGA Cys Pro Pro Gly His Ser Gly Ala Lys Cys Gln Glu Val Ser Gly Arg 850 855 860	2953
CCT TGC ATC ACC ATG GGG AGT GTG ATA CCA GAT GGG GCC AAA TGG GAT Pro Cys Ile Thr Met Gly Ser Val Ile Pro Asp Gly Ala Lys Trp Asp 865 870 875	3001
GAT GAC TGT AAT ACC TGC CAG TGC CTG AAT GGA CCG ATC GCC TGC TCA Asp Asp Cys Asn Thr Cys Gln Cys Leu Asn Gly Arg Ile Ala Cys Ser 880 885 890	3049
AAG GTC TGG TGT GGC CCT CGA CCT TGC CTG CTC CAC AAA GGG CAC AGC Lys Val Trp Cys Gly Pro Arg Pro Cys Leu Leu His Lys Gly His Ser 895 900 905	3097
GAG TGC CCC AGC GGG CAG AGC TGC ATC CCC ATC CTG GAC GAC CAG TGC Glu Cys Pro Ser Gly Gln Ser Cys Ile Pro Ile Leu Asp Asp Gln Cys 910 915 920 925	3145
TTC GTC CAC CCC TGC ACT GGT GTG GGC GAG TGT CGG TCT TCC AGT CTC Phe Val His Pro Cys Thr Gly Val Gly Glu Cys Arg Ser Ser Ser Leu 930 935 940	3193
CAG CCG GTG AAG ACA AAG TGC ACC TCT GAC TCC TAT TAC CAG GAT AAC Gln Pro Val Lys Thr Lys Cys Thr Ser Asp Ser Tyr Tyr Gln Asp Asn 945 950 955	3241
TGT GCG AAC ATC ACA TTT ACC TTT AAC AAG GAG ATG ATG TCA CCA GGT Cys Ala Asn Ile Thr Phe Thr Phe Asn Lys Glu Met Met Ser Pro Gly 960 965 970	3289
CTT ACT ACG GAG CAC ATT TGC AGT GAA TTG AGG AAT TTG AAT ATT TTG Leu Thr Thr Glu His Ile Cys Ser Glu Leu Arg Asn Leu Asn Ile Leu 975 980 985	3337
AAG AAT GTT TCC GCT GAA TAT TCA ATC TAC ATC GCT TGC GAG CCT TCC Lys Asn Val Ser Ala Glu Tyr Ser Ile Tyr Ile Ala Cys Glu Pro Ser 990 995 1000 1005	3385
CCT TCA GCG AAC AAT GAA ATA CAT GTG GCC ATT TCT GCT GAA GAT ATA Pro Ser Ala Asn Asn Glu Ile His Val Ala Ile Ser Ala Glu Asp Ile 1010 1015 1020	3433

CGG GAT GAT GGG AAC CCG ATC AAG GAA ATC ACT GAC AAA ATA ATC GAT Arg Asp Asp Gly Asn Pro Ile Lys Glu Ile Thr Asp Lys Ile Ile Asp 1025 1030 1035	3481
CTT GTT ACT AAA CGT GAT GGA AAC AGC TCG CTG ATT GCT GCC GTT GAA Leu Val Thr Lys Arg Asp Gly Asn Ser Ser Leu Ile Ala Ala Val Glu 1040 1045 1050	3529
GAA GTA AGA GTT CAG AGG CGG CCT CTG AAG AAC AGA ACA GAT TTC CTT Glu Val Arg Val Gln Arg Arg Pro Leu Lys Asn Arg Thr Asp Phe Leu 1055 1060 1065	3577
GTT CCC TTG CTG AGC TCT GTC TTA ACT GTG GCT TGG ATC TGT TGC TTG Val Pro Leu Leu Ser Ser Val Leu Thr Val Ala Trp Ile Cys Cys Leu 1070 1075 1080 1085	3625
GTG ACG GCC TTC TAC TGG TGC CTG CGG AAG CGG CGG AAG CCG GGC AGC Val Thr Ala Phe Tyr Trp Cys Leu Arg Lys Arg Arg Lys Pro Gly Ser 1090 1095 1100	3673
CAC ACA CAC TCA GCC TCT GAG GAC AAC ACC ACC AAC AAC GTG CGG GAG His Thr His Ser Ala Ser Glu Asp Asn Thr Thr Asn Asn Val Arg Glu 1105 1110 1115	3721
CAG CTG AAC CAG ATC AAA AAC CCC ATT GAG AAA CAT GGG GCC AAC ACG Gln Leu Asn Gln Ile Lys Asn Pro Ile Glu Lys His Gly Ala Asn Thr 1120 1125 1130	3769
GTC CCC ATC AAG GAT TAC GAG AAC AAG AAC TCC AAA ATG TCT AAA ATA Val Pro Ile Lys Asp Tyr Glu Asn Lys Asn Ser Lys Met Ser Lys Ile 1135 1140 1145	3817
AGG ACA CAC AAT TCT GAA GTA GAA GAG GAC GAC ATG GAC AAA CAC CAG Arg Thr His Asn Ser Glu Val Glu Glu Asp Asp Met Asp Lys His Gln 1150 1155 1160 1165	3865
CAG AAA GCC CGG TTT GCC AAG CAG CCG GCG TAC ACG CTG GTA GAC AGA Gln Lys Ala Arg Phe Ala Lys Gln Pro Ala Tyr Thr Leu Val Asp Arg 1170 1175 1180	3913
GAA GAG AAG CCC CCC AAC GGC ACG CCG ACA AAA CAC CCA AAC TGG ACA Glu Glu Lys Pro Pro Asn Gly Thr Pro Thr Lys His Pro Asn Trp Thr 1185 1190 1195	3961
AAC AAA CAG GAC AAC AGA GAC TTG GAA AGT GCC CAG AGC TTA AAC CGA Asn Lys Gln Asp Asn Arg Asp Leu Glu Ser Ala Gln Ser Leu Asn Arg 1200 1205 1210	4009
ATG GAG TAC ATC GTA TAG CAGACCGCGG GCACTGCCGC CGCTAGGTAG Met Glu Tyr Ile Val 1215	4057
AGTCTGAGGG CTTGTAGTTC TTAAACTGT CGTGTACATC TCGAGTCTGA GGCCGTTGCT	4117
GACTTAGAAT CCCTGTGTTA ATTTAGTTTG ACAAGCTGGC TTACACTGGC AATGGTAGTT	4177
CTGTGGTTGG CTGGGAAATC GAGTGGCGCA TCTCACAGCT ATGCAAAAAG CTAGTCAACA	4237
GTACCCCTGG TTGTGTGTCC CCTTGCAGCC GACACGGTCT CGGATCAGGC TCCCAGGAGC	4297
TGCCCAGCCC CCTGGTACTT TGAGCTCCCA CTTCTGCCAG ATGTCTAATG GTGATGCAGT	4357
CTTAGATCAT AGTTTTATTT ATATTTATTG ACTCTTGAGT TGTTTTTGTA TATTGGTTTT	4417
ATGATGACGT ACAAGTAGTT CTGTATTTGA AAGTGCCTTT GCAGCTCAGA ACCACAGCAA	4477
CGATCACAAA TGACTTTATT ATTTATTTTT TTTAATTGTA TTTTTGTTGT TGGGGGAGGG	4537

GAGACTTTGA	TGTCAGCAGT	TGCTGGTAAA	ATGAAGAATT	TAAAGAAAAA	ATGTCCAAAA	4597
GTAGAACTTT	GTATAGTTAT	GTAAATAATT	CTTTTTTATT	AATCACTGTG	TATATTTGAT	4657
TTATTAACCT	AATAATCAAG	AGCCTTAAAA	CATCATTCTT	TTTTATTTAT	ATGTATGTGT	4717
TTAGAATTGA	AGGTTTTTGA	TAGCATTGTA	AGCGTATGGC	TTTATTTTTT	TGAACTCTTC	4777
TCATTACTTG	TTGCCTATAA	GCCAAAAAGG	AAAGGGTGTT	TTGAAAATAG	TTTATTTTAA	4837
AACAATAGGA	TGGGCTACAC	GTACATAGGT	AAATAATAGC	ACCGTACTGG	TTATGATGAT	4897
GAAAATAACT	GGAAACTTGA	AAGCTTGTGG	TAATGGCAGA	TAAAGATGGT	TCACCTGGGA	4957
AATTAAAACT	TGAATGGTTG	TACAGAAAAG	CACAGAGTGG	AATGCACATC	AATGACAGTA	5017
AGGGAGTTAG	TTCTAGGAAC	AGCTCCTGAA	CAGTAAGATT	CCCGCAATAG	TCTCCGCCTC	5077
GTTCGTCTAT	GGTATGCATC	CCATTCAATT	TCTTCTTCTG	ATTATTGTCA	TCTTTCCCTT	5137
TGCCAAATGG	GCAGTTATTG	TTTCAGGGAG	AGAAGCTGCT	CATTGGCCAA	TCATTCTGGT	5197
GTGCAGTGCT	CCATCGGATT	CTACATGTCC	AACAAGGCAT	GTCTGGATGA	TGCAATGTCT	5257
GTCTGACCCC	CGGAATTCCG	TGCAGAGACA	ACATTCTAGA	CAGATATACA	CTTTTTATTA	5317
TTAACAAACT	TTGGCCACAA	CCTTTGATGT	ATAAATTGCC	GGATTTCCCC	AGTCCTTTCA	5377
TTGTGGCTTT	GGACAGGAGC	AGGCTCACTT	GTCTGCTTCA	GGCTGCCTTT	CTCTTGGGTT	5437
GCACCTCAGT	TCTTACTTAT	TTATTTATTT	TGAGTGGAGC	ATAGGGGCCT	CTTCCAAAAT	5497
GGGTAGAGCT	CAGGGGCTTT	CTTATTGAAA	TGGTCACATG	ATAAAAACGG	GCTGAAAAAG	5557
GAGAGTTCCA	GGAGAAAAGC	CCAGAAAAGG	CCCCTCCTCA	GAAGACAGCC	TTTAAGCCTC	5617
TTGCTTACTG	AAGGAAGCCC	CACCTTCTAG	CACTGAGGCC	GGGTCTGATC	TTCCAGAGGA	5677
GTTGGAGGAG	TCCATGAGAA	TGGCCACCAT	TCTTGCTTGC	TGCTGCTGAT	GTTGCAGTTT	5737
TGAGAGAACA	GCGGGATCCT	TGTTGTCCTC	TAGAGACTTG	AGTCTGTCAC	TGACATTTTT	5797
TCAGTTCCTT	TGCTCATAGA	CCATACGAGG	AATTAGTGAT	GTGTCAGTTG	AGAGTTCACA	5857
ATCTCATTGT	TCATTTAATT	CACTTTAAAG	TTGTCAATTT	CTGTGTGAGT	AACCTGTAAA	5917
AGACACCTTT	CCAGAAGAGT	TTTGCCGTCT	GTTTGAAAAA	AAAATCTTTA	TAACTTTTCC	5977
TAAGTATCTG	GATTTGGATT	CCTTATTTGG	AGAGAAAATG	TACCCTGTCT	CCACCAAAAA	6037
TACAAAAATT	AGCCAGGCTT	GGTGGTGCAC	ACCGGTAATC	CCAGCAACTC	TGGAGACTAA	6097
GGCAGGAAGA	ATCGCTTGAC	CCAGGAGGGT	CGAGGCTACA	ATGAGTTGAA	ACCGCGCCAC	6157
TGCACTCCAG	CCTGGGCGAC	AGTGCGAGGC	CCTGTCTCAA	AAATAAAATA	AAATAAATAA	6217
ATAAATTAGC	CAGATACTGT	GTGCACGCCT	GCAGTCCCAG	CTATTCTGGA	AGCTGAGGTG	6277
GGAAGATGGT	TAAGCCTGAG	AGGACAAAGC	TGCAGTGAGT	CATGTTTGCA	TCACTGCACT	6337
CCAGCCTGGG	TGACAGAGCA	AGACCCTGTC	TAAAAACAA	AAACAGGCCG	GGTGTGGTGG	6397
CTCATGCCTG	CCATCCCAGT	GCTTTGGGAG	GCAGAGGTTG	GCATAATCCC	AGCGCTCTGG	6457
GAATTCC						6464

## (2) INFORMATION FOR SEQ ID NO:2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1219 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```

Met Arg Ser Pro Arg Thr Arg Gly Arg Ser Gly Arg Pro Leu Ser Leu
 1           5           10           15
Leu Leu Ala Leu Leu Cys Ala Leu Arg Ala Lys Val Cys Gly Ala Ser
          20           25           30
Gly Gln Phe Glu Leu Glu Ile Leu Ser Met Gln Asn Val Asn Gly Glu
          35           40           45
Leu Gln Asn Gly Asn Cys Cys Gly Gly Ala Arg Asn Pro Gly Asp Arg
          50           55           60
Lys Cys Thr Arg Asp Glu Cys Asp Thr Tyr Phe Lys Val Cys Leu Lys
          65           70           75           80
Glu Tyr Gln Ser Arg Val Thr Ala Gly Gly Pro Cys Ser Phe Gly Ser
          85           90           95
Gly Ser Thr Pro Val Ile Gly Gly Asn Thr Phe Asn Leu Lys Ala Ser
          100          105          110
Arg Gly Asn Asp Pro Asn Arg Ile Val Leu Pro Phe Ser Phe Ala Trp
          115          120          125
Pro Arg Ser Tyr Thr Leu Leu Val Glu Ala Trp Asp Ser Ser Asn Asp
          130          135          140
Thr Val Gln Pro Asp Ser Ile Ile Glu Lys Ala Ser His Ser Gly Met
          145          150          155          160
Ile Asn Pro Ser Arg Gln Trp Gln Thr Leu Lys Gln Asn Thr Gly Val
          165          170          175
Ala His Phe Glu Tyr Gln Ile Arg Val Thr Cys Asp Asp Tyr Tyr Tyr
          180          185          190
Gly Phe Gly Cys Asn Lys Phe Cys Arg Pro Arg Asp Asp Phe Phe Gly
          195          200          205
His Tyr Ala Cys Asp Gln Asn Gly Asn Lys Thr Cys Met Glu Gly Trp
          210          215          220
Met Gly Pro Glu Cys Asn Arg Ala Ile Cys Arg Gln Gly Cys Ser Pro
          225          230          235          240
Lys His Gly Ser Cys Lys Leu Pro Gly Asp Cys Arg Cys Gln Tyr Gly
          245          250          255
Trp Gln Gly Leu Tyr Cys Asp Lys Cys Ile Pro His Pro Gly Cys Val
          260          265          270
His Gly Ile Cys Asn Glu Pro Trp Gln Cys Leu Cys Glu Thr Asn Trp
          275          280          285
Gly Gly Gln Leu Cys Asp Lys Asp Leu Asn Tyr Cys Gly Thr His Gln
          290          295          300

```

Pro Cys Leu Asn Gly Gly Thr Cys Ser Asn Thr Gly Pro Asp Lys Tyr  
 305 310 315 320  
 Gln Cys Ser Cys Pro Glu Gly Tyr Ser Gly Pro Asn Cys Glu Ile Ala  
 325 330 335  
 Glu His Ala Cys Leu Ser Asp Pro Cys His Asn Arg Gly Ser Cys Lys  
 340 345 350  
 Glu Thr Ser Leu Gly Phe Glu Cys Glu Cys Ser Pro Gly Trp Thr Gly  
 355 360 365  
 Pro Thr Cys Ser Thr Asn Ile Asp Asp Cys Ser Pro Asn Asn Cys Ser  
 370 375 380  
 His Gly Gly Thr Cys Gln Asp Leu Val Asn Gly Phe Lys Cys Val Cys  
 385 390 395 400  
 Pro Pro Gln Trp Thr Gly Lys Thr Cys Gln Leu Asp Ala Asn Glu Cys  
 405 410 415  
 Glu Ala Lys Pro Cys Val Asn Ala Lys Ser Cys Lys Asn Leu Ile Ala  
 420 425 430  
 Ser Tyr Tyr Cys Asp Cys Leu Pro Gly Trp Met Gly Gln Asn Cys Asp  
 435 440 445  
 Ile Asn Ile Asn Asp Cys Leu Gly Gln Cys Gln Asn Asp Ala Ser Cys  
 450 455 460  
 Arg Asp Leu Val Asn Gly Tyr Arg Cys Ile Cys Pro Pro Gly Tyr Ala  
 465 470 475 480  
 Gly Asp His Cys Glu Arg Asp Ile Asp Glu Cys Ala Ser Asn Pro Cys  
 485 490 495  
 Leu Asn Gly Gly His Cys Gln Asn Glu Ile Asn Arg Phe Gln Cys Leu  
 500 505 510  
 Cys Pro Thr Gly Phe Ser Gly Asn Leu Cys Gln Leu Asp Ile Asp Tyr  
 515 520 525  
 Cys Glu Pro Asn Pro Cys Gln Asn Gly Ala Gln Cys Tyr Asn Arg Ala  
 530 535 540  
 Ser Asp Tyr Phe Cys Lys Cys Pro Glu Asp Tyr Glu Gly Lys Asn Cys  
 545 550 555 560  
 Ser His Leu Lys Asp His Cys Arg Thr Thr Pro Cys Glu Val Ile Asp  
 565 570 575  
 Ser Cys Thr Val Ala Met Ala Ser Asn Asp Thr Pro Glu Gly Val Arg  
 580 585 590  
 Tyr Ile Ser Ser Asn Val Cys Gly Pro His Gly Lys Cys Lys Ser Gln  
 595 600 605  
 Ser Gly Gly Lys Phe Thr Cys Asp Cys Asn Lys Gly Phe Thr Gly Thr  
 610 615 620  
 Tyr Cys His Glu Asn Ile Asn Asp Cys Glu Ser Asn Pro Cys Arg Asn  
 625 630 635 640  
 Gly Gly Thr Cys Ile Asp Gly Val Asn Ser Tyr Lys Cys Ile Cys Ser  
 645 650 655  
 Asp Gly Trp Glu Gly Ala Tyr Cys Glu Thr Asn Ile Asn Asp Cys Ser

660										665					670				
Gln	Asn	Pro	Cys	His	Asn	Gly	Gly	Thr	Cys	Arg	Asp	Leu	Val	Asn	Asp				
		675					680					685							
Phe	Tyr	Cys	Asp	Cys	Lys	Asn	Gly	Trp	Lys	Gly	Lys	Thr	Cys	His	Ser				
	690					695					700								
Arg	Asp	Ser	Gln	Cys	Asp	Glu	Ala	Thr	Cys	Asn	Asn	Gly	Gly	Thr	Cys				
705					710					715								720	
Tyr	Asp	Glu	Gly	Asp	Ala	Phe	Lys	Cys	Met	Cys	Pro	Gly	Gly	Trp	Glu				
				725					730					735					
Gly	Thr	Thr	Cys	Asn	Ile	Ala	Arg	Asn	Ser	Ser	Cys	Leu	Pro	Asn	Pro				
			740					745						750					
Cys	His	Asn	Gly	Gly	Thr	Cys	Val	Val	Asn	Gly	Glu	Ser	Phe	Thr	Cys				
		755					760					765							
Val	Cys	Lys	Glu	Gly	Trp	Glu	Gly	Pro	Ile	Cys	Ala	Gln	Asn	Thr	Asn				
	770					775					780								
Asp	Cys	Ser	Pro	His	Pro	Cys	Tyr	Asn	Ser	Gly	Thr	Cys	Val	Asp	Gly				
785					790					795					800				
Asp	Asn	Trp	Tyr	Arg	Cys	Glu	Cys	Ala	Pro	Gly	Phe	Ala	Gly	Pro	Asp				
				805					810					815					
Cys	Arg	Ile	Asn	Ile	Asn	Glu	Cys	Gln	Ser	Ser	Pro	Cys	Ala	Phe	Gly				
			820					825					830						
Ala	Thr	Cys	Val	Asp	Glu	Ile	Asn	Gly	Tyr	Arg	Cys	Val	Cys	Pro	Pro				
		835					840					845							
Gly	His	Ser	Gly	Ala	Lys	Cys	Gln	Glu	Val	Ser	Gly	Arg	Pro	Cys	Ile				
	850				855						860								
Thr	Met	Gly	Ser	Val	Ile	Pro	Asp	Gly	Ala	Lys	Trp	Asp	Asp	Asp	Cys				
865					870					875					880				
Asn	Thr	Cys	Gln	Cys	Leu	Asn	Gly	Arg	Ile	Ala	Cys	Ser	Lys	Val	Trp				
			885						890					895					
Cys	Gly	Pro	Arg	Pro	Cys	Leu	Leu	His	Lys	Gly	His	Ser	Glu	Cys	Pro				
			900					905					910						
Ser	Gly	Gln	Ser	Cys	Ile	Pro	Ile	Leu	Asp	Asp	Gln	Cys	Phe	Val	His				
		915				920						925							
Pro	Cys	Thr	Gly	Val	Gly	Glu	Cys	Arg	Ser	Ser	Ser	Leu	Gln	Pro	Val				
	930					935					940								
Lys	Thr	Lys	Cys	Thr	Ser	Asp	Ser	Tyr	Tyr	Gln	Asp	Asn	Cys	Ala	Asn				
945					950					955					960				
Ile	Thr	Phe	Thr	Phe	Asn	Lys	Glu	Met	Met	Ser	Pro	Gly	Leu	Thr	Thr				
				965					970					975					
Glu	His	Ile	Cys	Ser	Glu	Leu	Arg	Asn	Leu	Asn	Ile	Leu	Lys	Asn	Val				
			980					985					990						
Ser	Ala	Glu	Tyr	Ser	Ile	Tyr	Ile	Ala	Cys	Glu	Pro	Ser	Pro	Ser	Ala				
		995					1000					1005							
Asn	Asn	Glu	Ile	His	Val	Ala	Ile	Ser	Ala	Glu	Asp	Ile	Arg	Asp	Asp				
	1010					1015					1020								



Gly Asn Pro Ile Lys Glu Ile Thr Asp Lys Ile Ile Asp Leu Val Thr  
 1025 1030 1035 1040  
 Lys Arg Asp Gly Asn Ser Ser Leu Ile Ala Ala Val Glu Glu Val Arg  
 1045 1050 1055  
 Val Gln Arg Arg Pro Leu Lys Asn Arg Thr Asp Phe Leu Val Pro Leu  
 1060 1065 1070  
 Leu Ser Ser Val Leu Thr Val Ala Trp Ile Cys Cys Leu Val Thr Ala  
 1075 1080 1085  
 Phe Tyr Trp Cys Leu Arg Lys Arg Arg Lys Pro Gly Ser His Thr His  
 1090 1095 1100  
 Ser Ala Ser Glu Asp Asn Thr Thr Asn Asn Val Arg Glu Gln Leu Asn  
 1105 1110 1115 1120  
 Gln Ile Lys Asn Pro Ile Glu Lys His Gly Ala Asn Thr Val Pro Ile  
 1125 1130 1135  
 Lys Asp Tyr Glu Asn Lys Asn Ser Lys Met Ser Lys Ile Arg Thr His  
 1140 1145 1150  
 Asn Ser Glu Val Glu Glu Asp Asp Met Asp Lys His Gln Gln Lys Ala  
 1155 1160 1165  
 Arg Phe Ala Lys Gln Pro Ala Tyr Thr Leu Val Asp Arg Glu Glu Lys  
 1170 1175 1180  
 Pro Pro Asn Gly Thr Pro Thr Lys His Pro Asn Trp Thr Asn Lys Gln  
 1185 1190 1195 1200  
 Asp Asn Arg Asp Leu Glu Ser Ala Gln Ser Leu Asn Arg Met Glu Tyr  
 1205 1210 1215  
 Ile Val

## (2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 4483 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 332..4483

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GGCCGGGGCC GGGCGGGCGG GTCGCGGGGG CAATGCGGGC GCAGGGCCGG GGGCGCCTTC	60
CCCGGCGGCT GCTGCTGCTG CTGGCGCTCT GGGTGCAGGC GGC <sup>1</sup> GCGGCC ATGGGCTATT	120
TCGAGCTGCA GCTGAGCGCG CTGCGGAACG TGAACGGGGA GCTGCTGAGC GGCGCCTGCT	180
GTGACGGCGA CGGCCGACA ACGCGCGCGG GGGGCTGCGG CCACGACGAG TCGACACCG	240
CTCCTTTACC CTCATCGTGG AGGCCTGGGA CTGGGACAAC GATACCACCC CGAATGAGGA	300

GCTGCTGATC	GAGCGAGTGT	CGCATGCCGG	C	ATG	ATC	AAC	CCG	GAG	GAC	CGC		352				
				Met	Ile	Asn	Pro	Glu	Asp	Arg						
				1				5								
TGG	AAG	AGC	CTG	CAC	TTC	AGC	GGC	CAC	GTG	GCG	CAC	CTG	GAG	CTG	CAG	400
Trp	Lys	Ser	Leu	His	Phe	Ser	Gly	His	Val	Ala	His	Leu	Glu	Leu	Gln	
		10					15					20				
ATC	CGC	GTG	CGC	TGC	GAC	GAG	AAC	TAC	TAC	AGC	GCC	ACT	TGC	AAC	AAG	448
Ile	Arg	Val	Arg	Cys	Asp	Glu	Asn	Tyr	Tyr	Ser	Ala	Thr	Cys	Asn	Lys	
	25					30					35					
TTC	TGC	CGG	CCC	CGC	AAT	GAC	TTT	TTC	GGC	CAC	TAC	ACC	TGC	GAC	CAG	496
Phe	Cys	Arg	Pro	Arg	Asn	Asp	Phe	Phe	Gly	His	Tyr	Thr	Cys	Asp	Gln	
	40				45					50					55	
TAC	GGC	AAC	AAG	GCC	TGC	ATG	GAC	GGC	TGG	ATG	GGC	AAG	GAG	TGC	AAG	544
Tyr	Gly	Asn	Lys	Ala	Cys	Met	Asp	Gly	Trp	Met	Gly	Lys	Glu	Cys	Lys	
				60					65					70		
GAA	GCT	GTG	TGT	AAA	CAA	GGG	TGT	AAT	TTG	CTC	CAC	GGG	GGA	TGC	ACC	592
Glu	Ala	Val	Cys	Lys	Gln	Gly	Cys	Asn	Leu	Leu	His	Gly	Gly	Cys	Thr	
			75					80					85			
GTG	CCT	GGG	GAG	TGC	AGG	TGC	AGC	TAC	GGC	TGG	CAA	GGG	AGG	TTC	TGC	640
Val	Pro	Gly	Glu	Cys	Arg	Cys	Ser	Tyr	Gly	Trp	Gln	Gly	Arg	Phe	Cys	
		90					95					100				
GAT	GAG	TGT	GTC	CCC	TAC	CCC	GGC	TGC	GTG	CAT	GGC	AGT	TGT	GTG	GAG	688
Asp	Glu	Cys	Val	Pro	Tyr	Pro	Gly	Cys	Val	His	Gly	Ser	Cys	Val	Glu	
	105					110					115					
CCC	TGG	CAG	TGC	AAC	TGT	GAG	ACC	AAC	TGG	GGC	GGC	CTG	CTC	TGT	GAC	736
Pro	Trp	Gln	Cys	Asn	Cys	Glu	Thr	Asn	Trp	Gly	Gly	Leu	Leu	Cys	Asp	
	120				125					130					135	
AAA	GAC	CTG	AAC	TAC	TGT	GGC	AGC	CAC	CAC	CCC	TGC	ACC	AAC	GGA	GGC	784
Lys	Asp	Leu	Asn	Tyr	Cys	Gly	Ser	His	His	Pro	Cys	Thr	Asn	Gly	Gly	
				140					145					150		
ACG	TGC	ATC	AAC	GCC	GAG	CCT	GAC	CAG	TAC	CGC	TGC	ACC	TGC	CCT	GAC	832
Thr	Cys	Ile	Asn	Ala	Glu	Pro	Asp	Gln	Tyr	Arg	Cys	Thr	Cys	Pro	Asp	
			155					160					165			
GGC	TAC	TCG	GGC	AGG	AAC	TGT	GAG	AAG	GCT	GAG	CAC	GCC	TGC	ACC	TCC	880
Gly	Tyr	Ser	Gly	Arg	Asn	Cys	Glu	Lys	Ala	Glu	His	Ala	Cys	Thr	Ser	
		170					175					180				
AAC	CCG	TGT	GCC	AAC	GGG	GGC	TCT	TGC	CAT	GAG	GTG	CCG	TCC	GGC	TTC	928
Asn	Pro	Cys	Ala	Asn	Gly	Gly	Ser	Cys	His	Glu	Val	Pro	Ser	Gly	Phe	
	185					190					195					
GAA	TGC	CAC	TGC	CCA	TCG	GGC	TGG	AGC	GGG	CCC	ACC	TGT	GCC	CTT	GAC	976
Glu	Cys	His	Cys	Pro	Ser	Gly	Trp	Ser	Gly	Pro	Thr	Cys	Ala	Leu	Asp	
	200				205					210				215		
ATC	GAT	GAG	TGT	GCT	TCG	AAC	CCG	TGT	GCG	GCC	GGT	GGC	ACC	TGT	GTG	1024
Ile	Asp	Glu	Cys	Ala	Ser	Asn	Pro	Cys	Ala	Ala	Gly	Gly	Thr	Cys	Val	
				220					225					230		
GAC	CAG	GTG	GAC	GGC	TTT	GAG	TGC	ATC	TGC	CCC	GAG	CAG	TGG	GTG	GGG	1072
Asp	Gln	Val	Asp	Gly	Phe	Glu	Cys	Ile	Cys	Pro	Glu	Gln	Trp	Val	Gly	
			235					240					245			
GCC	ACC	TGC	CAG	CTG	GAC	GCC	AAT	GAG	TGT	GAA	GGG	AAG	CCA	TGC	CTT	1120
Ala	Thr	Cys	Gln	Leu	Asp	Ala	Asn	Glu	Cys	Glu	Gly	Lys	Pro	Cys	Leu	
		250					255					260				

AAC GCT TTT TCT TGC AAA AAC CTG ATT GGC GGC TAT TAC TGT GAT TGC Asn Ala Phe Ser Cys Lys Asn Leu Ile Gly Gly Tyr Tyr Cys Asp Cys 265 270 275	1168
ATC CCG GGC TGG AAG GGC ATC AAC TGC CAT ATC AAC GTC AAC GAC TGT Ile Pro Gly Trp Lys Gly Ile Asn Cys His Ile Asn Val Asn Asp Cys 280 285 290 295	1216
CGC GGG CAG TGT CAG CAT GGG GGC ACC TGC AAG GAC CTG GTG AAC GGG Arg Gly Gln Cys Gln His Gly Gly Thr Cys Lys Asp Leu Val Asn Gly 300 305 310	1264
TAC CAG TGT GTG TGC CCA CGG GGC TTC GGA GGC CGG CAT TGC GAG CTG Tyr Gln Cys Val Cys Pro Arg Gly Phe Gly Gly Arg His Cys Glu Leu 315 320 325	1312
GAA CGA GAC AAG TGT GCC AGC AGC CCC TGC CAC AGC GGC GGC CTC TGC Glu Arg Asp Lys Cys Ala Ser Ser Pro Cys His Ser Gly Gly Leu Cys 330 335 340	1360
GAG GAC CTG GCC GAC GGC TTC CAC TGC CAC TGC CCC CAG GGC TTC TCC Glu Asp Leu Ala Asp Gly Phe His Cys His Cys Pro Gln Gly Phe Ser 345 350 355	1408
GGG CCT CTC TGT GAG GTG GAT GTC GAC CTT TGT GAG CCA AGC CCC TGC Gly Pro Leu Cys Glu Val Asp Val Asp Leu Cys Glu Pro Ser Pro Cys 360 365 370 375	1456
CGG AAC GGC GCT CGC TGC TAT AAC CTG GAG GGT GAC TAT TAC TGC GCC Arg Asn Gly Ala Arg Cys Tyr Asn Leu Glu Gly Asp Tyr Tyr Cys Ala 380 385 390	1504
TGC CCT GAT GAC TTT GGT GGC AAG AAC TGC TCC GTG CCC CGC GAG CCG Cys Pro Asp Asp Phe Gly Gly Lys Asn Cys Ser Val Pro Arg Glu Pro 395 400 405	1552
TGC CCT GGC GGG GCC TGC AGA GTG ATC GAT GGC TGC GGG TCA GAC GCG Cys Pro Gly Gly Ala Cys Arg Val Ile Asp Gly Cys Gly Ser Asp Ala 410 415 420	1600
GGG CCT GGG ATG CCT GGC ACA GCA GCC TCC GGC GTG TGT GGC CCC CAT Gly Pro Gly Met Pro Gly Thr Ala Ala Ser Gly Val Cys Gly Pro His 425 430 435	1648
GGA CGC TGC GTC AGC CAG CCA GGG GGC AAC TTT TCC TGC ATC TGT GAC Gly Arg Cys Val Ser Gln Pro Gly Gly Asn Phe Ser Cys Ile Cys Asp 440 445 450 455	1696
AGT GGC TTT ACT GGC ACC TAC TGC CAT GAG AAC ATT GAC GAC TGC CTG Ser Gly Phe Thr Gly Thr Tyr Cys His Glu Asn Ile Asp Asp Cys Leu 460 465 470	1744
GGC CAG CCC TGC CGC AAT GGG GGC ACA TGC ATC GAT GAG GTG GAC GCC Gly Gln Pro Cys Arg Asn Gly Gly Thr Cys Ile Asp Glu Val Asp Ala 475 480 485	1792
TTC CGC TGC TTC TGC CCC AGC GGT TGG GAG GGC GAG CTC TGC GAC ACC Phe Arg Cys Phe Cys Pro Ser Gly Trp Glu Gly Glu Leu Cys Asp Thr 490 495 500	1840
AAT CCC AAC GAC TGC CTT CCC GAT CCC TGC CAC AGC CGC GGC CGC TGC Asn Pro Asn Asp Cys Leu Pro Asp Pro Cys His Ser Arg Gly Arg Cys 505 510 515	1888
TAC GAC CTG GTC AAT GAC TTC TAC TGT GCG TGC GAC GAC GGC TGG AAG Tyr Asp Leu Val Asn Asp Phe Tyr Cys Ala Cys Asp Asp Gly Trp Lys 520 525 530 535	1936

GGC AAG ACC TGC CAC TCA CGC GAG TTC CAG TGC GAT GCC TAC ACC TGC Gly Lys Thr Cys His Ser Arg Glu Phe Gln Cys Asp Ala Tyr Thr Cys 540 545 550	1984
AGC AAC GGT GGC ACC TGC TAC GAC AGC GGC GAC ACC TTC CGC TGC GCC Ser Asn Gly Gly Thr Cys Tyr Asp Ser Gly Asp Thr Phe Arg Cys Ala 555 560 565	2032
TGC CCC CCC GGC TGG AAG GGC AGC ACC TGC GCC GTC GCC AAG AAC AGC Cys Pro Pro Gly Trp Lys Gly Ser Thr Cys Ala Val Ala Lys Asn Ser 570 575 580	2080
AGC TGC CTG CCC AAC CCC TGT GTG AAT GGT GGC ACC TGC GTG GGC AGC Ser Cys Leu Pro Asn Pro Cys Val Asn Gly Gly Thr Cys Val Gly Ser 585 590 595	2128
GGG GCC TCC TTC TCC TGC ATC TGC CGG GAC GGC TGG GAG GGT CGT ACT Gly Ala Ser Phe Ser Cys Ile Cys Arg Asp Gly Trp Glu Gly Arg Thr 600 605 610 615	2176
TGC ACT CAC AAT ACC AAC GAC TGC AAC CCT CTG CCT TGC TAC AAT GGT Cys Thr His Asn Thr Asn Asp Cys Asn Pro Leu Pro Cys Tyr Asn Gly 620 625 630	2224
GGC ATC TGT GTT GAC GGC GTC AAC TGG TTC CGC TGC GAG TGT GCA CCT Gly Ile Cys Val Asp Gly Val Asn Trp Phe Arg Cys Glu Cys Ala Pro 635 640 645	2272
GGC TTC GCG GGG CCT GAC TGC CGC ATC AAC ATC GAC GAG TGC CAG TCC Gly Phe Ala Gly Pro Asp Cys Arg Ile Asn Ile Asp Glu Cys Gln Ser 650 655 660	2320
TCG CCC TGT GCC TAC GGG GCC ACG TGT GTG GAT GAG ATC AAC GGG TAT Ser Pro Cys Ala Tyr Gly Ala Thr Cys Val Asp Glu Ile Asn Gly Tyr 665 670 675	2368
CGC TGT AGC TGC CCA CCC GGC CGA GCC GGC CCC CGG TGC CAG GAA GTG Arg Cys Ser Cys Pro Pro Gly Arg Ala Gly Pro Arg Cys Gln Glu Val 680 685 690 695	2416
ATC GGG TTC GGG AGA TCC TGC TGG TCC CGG GGC ACT CCG TTC CCA CAC Ile Gly Phe Gly Arg Ser Cys Trp Ser Arg Gly Thr Pro Phe Pro His 700 705 710	2464
GGA AGC TCC TGG GTG GAA GAC TGC AAC AGC TGC CGC TGC CTG GAT GGC Gly Ser Ser Trp Val Glu Asp Cys Asn Ser Cys Arg Cys Leu Asp Gly 715 720 725	2512
CGC CGT GAC TGC AGC AAG GTG TGG TGC GGA TGG AAG CCT TGT CTG CTG Arg Arg Asp Cys Ser Lys Val Trp Cys Gly Trp Lys Pro Cys Leu Leu 730 735 740	2560
GCC GGC CAG CCC GAG GCC CTG AGC GCC CAG TGC CCA CTG GGG CAA AGG Ala Gly Gln Pro Glu Ala Leu Ser Ala Gln Cys Pro Leu Gly Gln Arg 745 750 755	2608
TGC CTG GAG AAG GCC CCA GGC CAG TGT CTG CGA CCA CCC TGT GAG GCC Cys Leu Glu Lys Ala Pro Gly Gln Cys Leu Arg Pro Pro Cys Glu Ala 760 765 770 775	2656
TGG GGG GAG TGC GGC GCA GAA GAG CCA CCG AGC ACC CCC TGC CTG CCA Trp Gly Glu Cys Gly Ala Glu Glu Pro Pro Ser Thr Pro Cys Leu Pro 780 785 790	2704
CGC TCC GGC CAC CTG GAC AAT AAC TGT GCC CGC CTC ACC TTG CAT TTC Arg Ser Gly His Leu Asp Asn Asn Cys Ala Arg Leu Thr Leu His Phe 795 800 805	2752

AAC CGT GAC CAC GTG CCC CAG GGC ACC ACG GTG GGC GCC ATT TGC TCC Asn Arg Asp His Val Pro Gln Gly Thr Thr Val Gly Ala Ile Cys Ser 810 815 820	2800
GGG ATC CGC TCC CTG CCA GCC ACA AGG GCT GTG GCA CGG GAC CGC CTG Gly Ile Arg Ser Leu Pro Ala Thr Arg Ala Val Ala Arg Asp Arg Leu 825 830 835	2848
CTG GTG TTG CTT TGC GAC CGG GCG TCC TCG GGG GCC AGT GCT GTG GAG Leu Val Leu Leu Cys Asp Arg Ala Ser Ser Gly Ala Ser Ala Val Glu 840 845 850 855	2896
GTG GCC GTG TCC TTC AGC CCT GCC AGG GAC CTG CCT GAC AGC AGC CTG Val Ala Val Ser Phe Ser Pro Ala Arg Asp Leu Pro Asp Ser Ser Leu 860 865 870	2944
ATC CAG GGC GCG GCC CAC GCC ATC GTG GCC GCC ATC ACC CAG CGG GGG Ile Gln Gly Ala Ala His Ala Ile Val Ala Ala Ile Thr Gln Arg Gly 875 880 885	2992
AAC AGC TCA CTG CTC CTG GCT GTC ACC GAG GTC AAG GTG GAG ACC GTT Asn Ser Ser Leu Leu Leu Ala Val Thr Glu Val Lys Val Glu Thr Val 890 895 900	3040
GTT ACG GGC GGC TCT TCC ACA GGT CTG CTG GTG CCT GTG CTG TGT GGT Val Thr Gly Gly Ser Ser Thr Gly Leu Leu Val Pro Val Leu Cys Gly 905 910 915	3088
GCC TTC AGC GTG CTG TGG CTG GCG TGC GTG GTC CTG TGC GTG TGG TGG Ala Phe Ser Val Leu Trp Leu Ala Cys Val Val Leu Cys Val Trp Trp 920 925 930 935	3136
ACA CGC AAG CGC AGG AAA GAG CGG GAG AGG AGC CGG CTG CCG CGG GAG Thr Arg Lys Arg Arg Lys Glu Arg Glu Arg Ser Arg Leu Pro Arg Glu 940 945 950	3184
GAG AGC GCC AAC AAC CAG TGG GCC CCG CTC AAC CCC ATC CGC AAC CCC Glu Ser Ala Asn Asn Gln Trp Ala Pro Leu Asn Pro Ile Arg Asn Pro 955 960 965	3232
ATT GAG CGG CCG GGG GGG CAC AAG GAC GTG CTC TAC CAG TGC AAG AAC Ile Glu Arg Pro Gly Gly His Lys Asp Val Leu Tyr Gln Cys Lys Asn 970 975 980	3280
TTC ACT CCA CCG CCG CGC AGG CGC TGC CCG GGC CGG CCG GCC ACG CGG Phe Thr Pro Pro Pro Arg Arg Arg Cys Pro Gly Arg Pro Ala Thr Arg 985 990 995	3328
CCG TCA GGG AGG ATG AGG AGG ACG AGG ATC TTG GCC GCG GTG AGG AGG Pro Ser Gly Arg Met Arg Arg Thr Arg Ile Leu Ala Ala Val Arg Arg 1000 1005 1010 1015	3376
ACT CCC TGG AGG CGG AGA AGT TCC TCT CAC ACA AAT TCA CCA AAG ATC Thr Pro Trp Arg Arg Arg Ser Ser His Thr Asn Ser Pro Lys Ile 1020 1025 1030	3424
CTG GCC GCT CGC CGG GGA GGC CGG CCC ACT GGG CCT CAG GCC CCA AAG Leu Ala Ala Arg Arg Gly Gly Arg Pro Thr Gly Pro Gln Ala Pro Lys 1035 1040 1045	3472
TGG ACA ACC GCG CGG TCA GGA GCA TCA ATG AGG CCC GCT ACG TCG GCA Trp Thr Thr Ala Arg Ser Gly Ala Ser Met Arg Pro Ala Thr Ser Ala 1050 1055 1060	3520
AGG GAA GTA GGG CGG CTG CAG CTG GGC CGG GAC CCA GGG CCC TCG GTG Arg Glu Val Gly Arg Leu Gln Leu Gly Arg Asp Pro Gly Pro Ser Val 1065 1070 1075	3568

GGA GCC ATG CCG TCT GCC GGA CCC GGA GGC CGA GGC CAT GTG CAT AGT Gly Ala Met Pro Ser Ala Gly Pro Gly Gly Arg Gly His Val His Ser 1080 1085 1090 1095	3616
TTC TTT ATT TTG TGT AAA AAA ACC ACC AAA AAC AAA AAC CAA ATG TTT Phe Phe Ile Leu Cys Lys Lys Thr Thr Lys Asn Lys Asn Gln Met Phe 1100 1105 1110	3664
ATT TTC TAC GTT TCT TTA ACC TTG TAT AAA TTA TTC AGT AAC TGT CAG Ile Phe Tyr Val Ser Leu Thr Leu Tyr Lys Leu Phe Ser Asn Cys Gln 1115 1120 1125	3712
GCT GAA AAC AAT GGA GTA TTC TCG GAT AGT TGC TAT TTT TGT AAA GTA Ala Glu Asn Asn Gly Val Phe Ser Asp Ser Cys Tyr Phe Cys Lys Val 1130 1135 1140	3760
GCC GTG CGT GGC ACT CGC TGT ATG AAA GGA GAG AGC AAA GGG TGT CTG Ala Val Arg Gly Thr Arg Cys Met Lys Gly Glu Ser Lys Gly Cys Leu 1145 1150 1155	3808
CGT CGT CAC CAA ATC GTC GCG TTT GTT ACC AGA GGT TGT GCA CTG TTT Arg Arg His Gln Ile Val Ala Phe Val Thr Arg Gly Cys Ala Leu Phe 1160 1165 1170 1175	3856
ACA GAA TCT TCC TTT TAT TCC TCA CTC GGG TTT CTC TGT GCT CCA GGC Thr Glu Ser Ser Phe Tyr Ser Ser Leu Gly Phe Leu Cys Ala Pro Gly 1180 1185 1190	3904
CAA AGT GCC GGT GAG ACC CAT GGC TGT GTT GGT GTG GCC CAT GGC TGT Gln Ser Ala Gly Glu Thr His Gly Cys Val Gly Val Ala His Gly Cys 1195 1200 1205	3952
TGG TGG GAC CCG TGG CTG ATG GTG TGG CCT GTG GCT GTC GGT GGG ACT Trp Trp Asp Pro Trp Leu Met Val Trp Pro Val Ala Val Gly Gly Thr 1210 1215 1220	4000
CGT GGC TGT CAA TGG GAC CTG TGG CTG TCG GTG GGA CCT ACG GTG GTC Arg Gly Cys Gln Trp Asp Leu Trp Leu Ser Val Gly Pro Thr Val Val 1225 1230 1235	4048
GGT GGG ACC CTG GTT ATT GAT GTG GCC CTG GCT GCC GGC ACG GCC CGT Gly Gly Thr Leu Val Ile Asp Val Ala Leu Ala Ala Gly Thr Ala Arg 1240 1245 1250 1255	4096
GGC TGT TG ACGCACCT GTGGTTGTTA GTGGGGCCCTG AGGTCATCGGC GTGGCCCAAG Gly Cys	4154
GCCGGCAGGT CAACCTCGCG CTTGCTGGCC AGTCCACCCT GCCTGCCGTCT GTGCTTCCTC	4214
CTGCCCAGAA CGCCCGCTCC AGCGATCTCT CCACTGTGCT TTCAGAAGTGC CCTTCCTGCT	4274
GCGCAGTTCT CCCATCCTGG GACGGCGGCA GTATTGAAGC TCGTGACAAGT GCCTTCACAC	4334
AGACCCCTCG CAACTGTCCA CGCGTGCCGT GGCACCAGGC GCTGCCCACCT GCCGGCCCCG	4394
GCCGCCCCCTC CTCGTGAAAG TGCATTTTTG TAAATGTGTA CATATTAAAGG AAGCACTCTG	4454
TATAAAAAAA AAAAACCGGA ATTCC	4483

## (2) INFORMATION FOR SEQ ID NO:4:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1384 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Ile Asn Pro Glu Asp Arg Trp Lys Ser Leu His Phe Ser Gly His  
 1 5 10 15  
 Val Ala His Leu Glu Leu Gln Ile Arg Val Arg Cys Asp Glu Asn Tyr  
 20 25 30  
 Tyr Ser Ala Thr Cys Asn Lys Phe Cys Arg Pro Arg Asn Asp Phe Phe  
 35 40 45  
 Gly His Tyr Thr Cys Asp Gln Tyr Gly Asn Lys Ala Cys Met Asp Gly  
 50 55 60  
 Trp Met Gly Lys Glu Cys Lys Glu Ala Val Cys Lys Gln Gly Cys Asn  
 65 70 75 80  
 Leu Leu His Gly Gly Cys Thr Val Pro Gly Glu Cys Arg Cys Ser Tyr  
 85 90 95  
 Gly Trp Gln Gly Arg Phe Cys Asp Glu Cys Val Pro Tyr Pro Gly Cys  
 100 105 110  
 Val His Gly Ser Cys Val Glu Pro Trp Gln Cys Asn Cys Glu Thr Asn  
 115 120 125  
 Trp Gly Gly Leu Leu Cys Asp Lys Asp Leu Asn Tyr Cys Gly Ser His  
 130 135 140  
 His Pro Cys Thr Asn Gly Gly Thr Cys Ile Asn Ala Glu Pro Asp Gln  
 145 150 155 160  
 Tyr Arg Cys Thr Cys Pro Asp Gly Tyr Ser Gly Arg Asn Cys Glu Lys  
 165 170 175  
 Ala Glu His Ala Cys Thr Ser Asn Pro Cys Ala Asn Gly Gly Ser Cys  
 180 185 190  
 His Glu Val Pro Ser Gly Phe Glu Cys His Cys Pro Ser Gly Trp Ser  
 195 200 205  
 Gly Pro Thr Cys Ala Leu Asp Ile Asp Glu Cys Ala Ser Asn Pro Cys  
 210 215 220  
 Ala Ala Gly Gly Thr Cys Val Asp Gln Val Asp Gly Phe Glu Cys Ile  
 225 230 235 240  
 Cys Pro Glu Gln Trp Val Gly Ala Thr Cys Gln Leu Asp Ala Asn Glu  
 245 250 255  
 Cys Glu Gly Lys Pro Cys Leu Asn Ala Phe Ser Cys Lys Asn Leu Ile  
 260 265 270  
 Gly Gly Tyr Tyr Cys Asp Cys Ile Pro Gly Trp Lys Gly Ile Asn Cys  
 275 280 285  
 His Ile Asn Val Asn Asp Cys Arg Gly Gln Cys Gln His Gly Gly Thr  
 290 295 300  
 Cys Lys Asp Leu Val Asn Gly Tyr Gln Cys Val Cys Pro Arg Gly Phe  
 305 310 315 320  
 Gly Gly Arg His Cys Glu Leu Glu Arg Asp Lys Cys Ala Ser Ser Pro  
 325 330 335

Cys His Ser Gly Gly Leu Cys Glu Asp Leu Ala Asp Gly Phe His Cys  
 340 345 350  
 His Cys Pro Gln Gly Phe Ser Gly Pro Leu Cys Glu Val Asp Val Asp  
 355 360 365  
 Leu Cys Glu Pro Ser Pro Cys Arg Asn Gly Ala Arg Cys Tyr Asn Leu  
 370 375 380  
 Glu Gly Asp Tyr Tyr Cys Ala Cys Pro Asp Asp Phe Gly Gly Lys Asn  
 385 390 395 400  
 Cys Ser Val Pro Arg Glu Pro Cys Pro Gly Gly Ala Cys Arg Val Ile  
 405 410 415  
 Asp Gly Cys Gly Ser Asp Ala Gly Pro Gly Met Pro Gly Thr Ala Ala  
 420 425 430  
 Ser Gly Val Cys Gly Pro His Gly Arg Cys Val Ser Gln Pro Gly Gly  
 435 440 445  
 Asn Phe Ser Cys Ile Cys Asp Ser Gly Phe Thr Gly Thr Tyr Cys His  
 450 455 460  
 Glu Asn Ile Asp Asp Cys Leu Gly Gln Pro Cys Arg Asn Gly Gly Thr  
 465 470 475 480  
 Cys Ile Asp Glu Val Asp Ala Phe Arg Cys Phe Cys Pro Ser Gly Trp  
 485 490 495  
 Glu Gly Glu Leu Cys Asp Thr Asn Pro Asn Asp Cys Leu Pro Asp Pro  
 500 505 510  
 Cys His Ser Arg Gly Arg Cys Tyr Asp Leu Val Asn Asp Phe Tyr Cys  
 515 520 525  
 Ala Cys Asp Asp Gly Trp Lys Gly Lys Thr Cys His Ser Arg Glu Phe  
 530 535 540  
 Gln Cys Asp Ala Tyr Thr Cys Ser Asn Gly Gly Thr Cys Tyr Asp Ser  
 545 550 555 560  
 Gly Asp Thr Phe Arg Cys Ala Cys Pro Pro Gly Trp Lys Gly Ser Thr  
 565 570 575  
 Cys Ala Val Ala Lys Asn Ser Ser Cys Leu Pro Asn Pro Cys Val Asn  
 580 585 590  
 Gly Gly Thr Cys Val Gly Ser Gly Ala Ser Phe Ser Cys Ile Cys Arg  
 595 600 605  
 Asp Gly Trp Glu Gly Arg Thr Cys Thr His Asn Thr Asn Asp Cys Asn  
 610 615 620  
 Pro Leu Pro Cys Tyr Asn Gly Gly Ile Cys Val Asp Gly Val Asn Trp  
 625 630 635 640  
 Phe Arg Cys Glu Cys Ala Pro Gly Phe Ala Gly Pro Asp Cys Arg Ile  
 645 650 655  
 Asn Ile Asp Glu Cys Gln Ser Ser Pro Cys Ala Tyr Gly Ala Thr Cys  
 660 665 670  
 Val Asp Glu Ile Asn Gly Tyr Arg Cys Ser Cys Pro Pro Gly Arg Ala  
 675 680 685  
 Gly Pro Arg Cys Gln Glu Val Ile Gly Phe Gly Arg Ser Cys Trp Ser



690					695					700					
Arg 705	Gly	Thr	Pro	Phe	Pro 710	His	Gly	Ser	Ser	Trp 715	Val	Glu	Asp	Cys	Asn 720
Ser	Cys	Arg	Cys	Leu 725	Asp	Gly	Arg	Arg	Asp 730	Cys	Ser	Lys	Val	Trp 735	Cys
Gly	Trp	Lys	Pro 740	Cys	Leu	Leu	Ala	Gly 745	Gln	Pro	Glu	Ala	Leu 750	Ser	Ala
Gln	Cys	Pro 755	Leu	Gly	Gln	Arg	Cys 760	Leu	Glu	Lys	Ala	Pro 765	Gly	Gln	Cys
Leu 770	Arg	Pro	Pro	Cys	Glu	Ala 775	Trp	Gly	Glu	Cys	Gly 780	Ala	Glu	Glu	Pro
Pro 785	Ser	Thr	Pro	Cys	Leu 790	Pro	Arg	Ser	Gly	His 795	Leu	Asp	Asn	Asn	Cys 800
Ala	Arg	Leu	Thr	Leu 805	His	Phe	Asn	Arg	Asp 810	His	Val	Pro	Gln	Gly 815	Thr
Thr	Val	Gly	Ala 820	Ile	Cys	Ser	Gly	Ile 825	Arg	Ser	Leu	Pro	Ala	Thr	Arg
Ala	Val	Ala 835	Arg	Asp	Arg	Leu	Leu 840	Val	Leu	Leu	Cys	Asp 845	Arg	Ala	Ser
Ser	Gly 850	Ala	Ser	Ala	Val	Glu 855	Val	Ala	Val	Ser	Phe	Ser	Pro	Ala	Arg
Asp 865	Leu	Pro	Asp	Ser	Ser	Leu	Ile	Gln	Gly	Ala 875	Ala	His	Ala	Ile	Val 880
Ala	Ala	Ile	Thr	Gln 885	Arg	Gly	Asn	Ser	Ser 890	Leu	Leu	Leu	Ala	Val 895	Thr
Glu	Val	Lys	Val 900	Glu	Thr	Val	Val	Thr 905	Gly	Gly	Ser	Ser	Thr	Gly 910	Leu
Leu	Val	Pro	Val	Leu	Cys	Gly	Ala 920	Phe	Ser	Val	Leu	Trp 925	Leu	Ala	Cys
Val 930	Val	Leu	Cys	Val	Trp	Trp	Thr	Arg	Lys	Arg	Arg 940	Lys	Glu	Arg	Glu
Arg 945	Ser	Arg	Leu	Pro	Arg 950	Glu	Glu	Ser	Ala	Asn 955	Asn	Gln	Trp	Ala	Pro 960
Leu	Asn	Pro	Ile	Arg 965	Asn	Pro	Ile	Glu	Arg 970	Pro	Gly	Gly	His	Lys 975	Asp
Val	Leu	Tyr	Gln 980	Cys	Lys	Asn	Phe	Thr 985	Pro	Pro	Pro	Arg	Arg	Arg	Cys
Pro	Gly	Arg 995	Pro	Ala	Thr	Arg	Pro	Ser 1000	Gly	Arg	Met	Arg 1005	Arg	Thr	Arg
Ile 1010	Leu	Ala	Ala	Val	Arg	Arg	Thr	Pro	Trp	Arg	Arg 1020	Arg	Ser	Ser	Ser
His 1025	Thr	Asn	Ser	Pro	Lys 1030	Ile	Leu	Ala	Ala	Arg 1035	Arg	Gly	Gly	Arg	Pro 1040
Thr	Gly	Pro	Gln	Ala 1045	Pro	Lys	Trp	Thr	Thr	Ala	Arg	Ser	Gly	Ala	Ser 1055

Met Arg Pro Ala Thr Ser Ala Arg Glu Val Gly Arg Leu Gln Leu Gly  
 1060 1065 1070

Arg Asp Pro Gly Pro Ser Val Gly Ala Met Pro Ser Ala Gly Pro Gly  
 1075 1080 1085

Gly Arg Gly His Val His Ser Phe Phe Ile Leu Cys Lys Lys Thr Thr  
 1090 1095 1100

Lys Asn Lys Asn Gln Met Phe Ile Phe Tyr Val Ser Leu Thr Leu Tyr  
 1105 1110 1115 1120

Lys Leu Phe Ser Asn Cys Gln Ala Glu Asn Asn Gly Val Phe Ser Asp  
 1125 1130 1135

Ser Cys Tyr Phe Cys Lys Val Ala Val Arg Gly Thr Arg Cys Met Lys  
 1140 1145 1150

Gly Glu Ser Lys Gly Cys Leu Arg Arg His Gln Ile Val Ala Phe Val  
 1155 1160 1165

Thr Arg Gly Cys Ala Leu Phe Thr Glu Ser Ser Phe Tyr Ser Ser Leu  
 1170 1175 1180

Gly Phe Leu Cys Ala Pro Gly Gln Ser Ala Gly Glu Thr His Gly Cys  
 1185 1190 1195 1200

Val Gly Val Ala His Gly Cys Trp Trp Asp Pro Trp Leu Met Val Trp  
 1205 1210 1215

Pro Val Ala Val Gly Gly Thr Arg Gly Cys Gln Trp Asp Leu Trp Leu  
 1220 1225 1230

Ser Val Gly Pro Thr Val Val Gly Gly Thr Leu Val Ile Asp Val Ala  
 1235 1240 1245

Leu Ala Ala Gly Thr Ala Arg Gly Cys  
 1250 1255

## (2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3582 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

- (ix) FEATURE:
- (A) NAME/KEY: CDS
  - (B) LOCATION: 1..3582

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CAG GTG GCG TCA GCA TCG GGA CAG TTC GAG CTG GAG ATC TTA TCC GTG	48
Gln Val Ala Ser Ala Ser Gly Gln Phe Glu Leu Glu Ile Leu Ser Val	
1 5 10 15	
CAG AAT GTG AAC GGC GTG CTG CAG AAC GGG AAC TGC TGC GAC GGC ACT	96
Gln Asn Val Asn Gly Val Leu Gln Asn Gly Asn Cys Cys Asp Gly Thr	
20 25 30	

CGA AAC CCC GGA GAT AAA AAG TGC ACC AGA GAT GAG TGT GAC ACC TAC Arg Asn Pro Gly Asp Lys Lys Cys Thr Arg Asp Glu Cys Asp Thr Tyr 35 40 45	144
TTT AAA GTT TGC CTG AAG GA TAC CAG TCG CGG GTC ACT GCT GGC GGC Phe Lys Val Cys Leu Lys Glu Tyr Gln Ser Arg Val Thr Ala Gly Gly 50 55 60	192
CCT TGC AGC TTC GGA TCC AAA TCC ACC CCT GTC ATC GGC GGG AAT ACC Pro Cys Ser Phe Gly Ser Lys Ser Thr Pro Val Ile Gly Gly Asn Thr 65 70 75 80	240
TTC AAT TTA AAG TAC AGC CGG AAT AAT GAA AAG AAC CGG ATT GTT ATC Phe Asn Leu Lys Tyr Ser Arg Asn Asn Glu Lys Asn Arg Ile Val Ile 85 90 95	288
CCT TTC ACG TTC GCC TGG CCG AGA TCC TAC ACG TTG CTT GTT GAG GCA Pro Phe Thr Phe Ala Trp Pro Arg Ser Tyr Thr Leu Leu Val Glu Ala 100 105 110	336
TGG GAT TAC AAT GAT AAC TCT ACT AAT CCC GAT CGC ATA ATT GAG AAG Trp Asp Tyr Asn Asp Asn Ser Thr Asn Pro Asp Arg Ile Ile Glu Lys 115 120 125	384
GCA TCC CAC TCT GGC ATG ATC AAT CCA AGC CGT CAG TGG CAG ACG TTG Ala Ser His Ser Gly Met Ile Asn Pro Ser Arg Gln Trp Gln Thr Leu 130 135 140	432
AAA CAT AAC ACA GGA GCT GCC CAC TTT GAG TAT CAA ATC CGT GTG ACT Lys His Asn Thr Gly Ala Ala His Phe Glu Tyr Gln Ile Arg Val Thr 145 150 155 160	480
TGC GCA GAA CAT TAC TAT GGC TTT GGA TGC AAC AAG TTT TGT CGA CCG Cys Ala Glu His Tyr Tyr Gly Phe Gly Cys Asn Lys Phe Cys Arg Pro 165 170 175	528
AGA GAT GAC TTC TTC ACT CAC CAT ACC TGT GAC CAG AAT GGC AAC AAA Arg Asp Asp Phe Phe Thr His His Thr Cys Asp Gln Asn Gly Asn Lys 180 185 190	576
ACC TGC TTG GAA GGC TGG ACG GGA CCA GAA TGC AAC AAA GCT ATT TGT Thr Cys Leu Glu Gly Trp Thr Gly Pro Glu Cys Asn Lys Ala Ile Cys 195 200 205	624
CGT CAG GGA TGT AGC CCC AAG CAT GGT TCT TGC ACA GTT CCA GGA GAG Arg Gln Gly Cys Ser Pro Lys His Gly Ser Cys Thr Val Pro Gly Glu 210 215 220	672
TGC AGG TGT CAG TAT GGA TGG CAA GGC CAG TAC TGT GAT AAG TGC ATT Cys Arg Cys Gln Tyr Gly Trp Gln Gly Gln Tyr Cys Asp Lys Cys Ile 225 230 235 240	720
CCA CAC CCG GGA TGT GTC CAT GGC ACT TGC ATT GAA CCA TGG CAG TGC Pro His Pro Gly Cys Val His Gly Thr Cys Ile Glu Pro Trp Gln Cys 245 250 255	768
CTC TGT GAA ACC AAC TGG GGT GGT CAG CTC TGT GAC AAA GAC CTG AAC Leu Cys Glu Thr Asn Trp Gly Gly Gln Leu Cys Asp Lys Asp Leu Asn 260 265 270	816
TAC TGT GGA ACC CAC CCA CCC TGT TTG AAT GGT GGT ACC TGC AGC AAC Tyr Cys Gly Thr His Pro Pro Cys Leu Asn Gly Gly Thr Cys Ser Asn 275 280 285	864
ACT GGC CCC GAT AAA TAC CAG TGT TCC TGC CCT GAG GGT TAC TCA GGA Thr Gly Pro Asp Lys Tyr Gln Cys Ser Cys Pro Glu Gly Tyr Ser Gly 290 295 300	912

CAG AAC TGT GAA ATA GCG GAG CAT GCG TGC CTC TCT GAT CCG TGC CAC Gln Asn Cys Glu Ile Ala Glu His Ala Cys Leu Ser Asp Pro Cys His 305 310 315 320	960
AAC GGA GGA AGC TGC CTA GAA ACG TCT ACA GGA TTT GAA TGT GTG TGT Asn Gly Gly Ser Cys Leu Glu Thr Ser Thr Gly Phe Glu Cys Val Cys 325 330 335	1008
GCA CCT GGC TGG GCT GGA CCA ACT TGC ACT GAT AAT ATT GAT GAT TGT Ala Pro Gly Trp Ala Gly Pro Thr Cys Thr Asp Asn Ile Asp Asp Cys 340 345 350	1056
TCT CCA AAT CCC TGT GGT CAT GGA GGA ACT TGC CAA GAT CTA GTT GAT Ser Pro Asn Pro Cys Gly His Gly Thr Cys Gln Asp Leu Val Asp 355 360 365	1104
GGA TTT AAG TGT ATT TGC CCA CCT CAG TGG ACT GGC AAA ACA TGC CAG Gly Phe Lys Cys Ile Cys Pro Gln Trp Thr Gly Lys Thr Cys Gln 370 375 380	1152
CTA GAT GCG AAT GAA TGT GAG GGC AAA CCC TGT GTC AAT GCC AAC TCC Leu Asp Ala Asn Glu Cys Glu Gly Lys Pro Cys Val Asn Ala Asn Ser 385 390 395 400	1200
TGC AGG AAC TTG ATT GGC AGC TAC TAT TGT GAC TGC ATT ACT GGC TGG Cys Arg Asn Leu Ile Gly Ser Tyr Tyr Cys Asp Cys Ile Thr Gly Trp 405 410 415	1248
TCT GGC CAC AAC TGT GAT ATA AAT ATT AAT GAT TGT CGT GGA CAA TGT Ser Gly His Asn Cys Asp Ile Asn Ile Asn Asp Cys Arg Gly Gln Cys 420 425 430	1296
CAG AAT GGA GGA TCC TGT CGG GAC TTG GTT AAT GGT TAT CGG TGC ATC Gln Asn Gly Gly Ser Cys Arg Asp Leu Val Asn Gly Tyr Arg Cys Ile 435 440 445	1344
TGT TCA CCT GGC TAT GCA GGA GAT CAC TGT GAG AAA GAC ATC AAT GAA Cys Ser Pro Gly Tyr Ala Gly Asp His Cys Glu Lys Asp Ile Asn Glu 450 455 460	1392
TGT GCA AGT AAC CCT TGC ATG AAT GGG GGT CAC TGC CAG GAT GAA ATC Cys Ala Ser Asn Pro Cys Met Asn Gly Gly His Cys Gln Asp Glu Ile 465 470 475 480	1440
AAT GGA TTC CAA TGT CTG TGT CCT GCT GGT TTC TCA GGA AAC CTC TGT Asn Gly Phe Gln Cys Leu Cys Pro Ala Gly Phe Ser Gly Asn Leu Cys 485 490 495	1488
CAG CTG GAT ATA GAC TAC TGT GAG CCA AAC CCT TGC CAG AAC GGT GCC Gln Leu Asp Ile Asp Tyr Cys Glu Pro Asn Pro Cys Gln Asn Gly Ala 500 505 510	1536
CAG TGC TTC AAT CTT GCT ATG GAC TAT TTC TGT AAC TGC CCT GAA GAT Gln Cys Phe Asn Leu Ala Met Asp Tyr Phe Cys Asn Cys Pro Glu Asp 515 520 525	1584
TAC GAA GGC AAG AAC TGC TCC CAC CTG AAA GAT CAC TGC CGC ACA ACT Tyr Glu Gly Lys Asn Cys Ser His Leu Lys Asp His Cys Arg Thr Thr 530 535 540	1632
CCT TGT GAA GTA ATC GAC AGC TGT ACA GTG GCA GTG GCT TCT AAC AGC Pro Cys Glu Val Ile Asp Ser Cys Thr Val Ala Val Ala Ser Asn Ser 545 550 555 560	1680
ACA CCA GAA GGA GTT CGT TAC ATT TCT TCA AAT GTC TGT GGT CCT CAT Thr Pro Glu Gly Val Arg Tyr Ile Ser Ser Asn Val Cys Gly Pro His 565 570 575	1728

GGA AAA TGC AAG AGC CAA GCA GGT GGA AAA TTC ACC TGT GAA TGC AAC	1776
Gly Lys Cys Lys Ser Gln Ala Gly Gly Lys Phe Thr Cys Glu Cys Asn	
580	
585	
590	
AAA GGA TTC ACT GGC ACC TAC TGT CAT GAG AAT ATC AAT GAC TGT GAG	1824
Lys Gly Phe Thr Gly Thr Tyr Cys His Glu Asn Ile Asn Asp Cys Glu	
595	
600	
605	
AGC AAC CCC TGT AAA AAT GGT GGC ACT TGT ATT GAC GGT GTA AAC TCC	1872
Ser Asn Pro Cys Lys Asn Gly Gly Thr Cys Ile Asp Gly Val Asn Ser	
610	
615	
620	
TAC AAA TGT ATT TGT AGT GAT GGA TGG GAA GGA ACA TAT TGT GAA ACA	1920
Tyr Lys Cys Ile Cys Ser Asp Gly Trp Glu Gly Thr Tyr Cys Glu Thr	
625	
630	
635	
640	
AAT ATT AAT GAC TGC AGT AAA AAC CCC TGC CAC AAT GGA GGA ACT TGC	1968
Asn Ile Asn Asp Cys Ser Lys Asn Pro Cys His Asn Gly Gly Thr Cys	
645	
650	
655	
CGA GAC TTG GTC AAT GAC TTC TTC TGT GAA TGT AAA AAT GGG TGG AAA	2016
Arg Asp Leu Val Asn Asp Phe Phe Cys Glu Cys Lys Asn Gly Trp Lys	
660	
665	
670	
GGA AAA ACT TGC CAC TCT CGT GAC AGC CAG TGT GAT GAG GCA ACA TGC	2064
Gly Lys Thr Cys His Ser Arg Asp Ser Gln Cys Asp Glu Ala Thr Cys	
675	
680	
685	
AAT AAT GGA GGA ACA TGT TAT GAT GAG GGG GAC ACT TTC AAG TGC ATG	2112
Asn Asn Gly Gly Thr Cys Tyr Asp Glu Gly Asp Thr Phe Lys Cys Met	
690	
695	
700	
TGT CCT GCA GGA TGG GAA GGA GCC ACT TGT AAT ATA GCA AGG AAC AGC	2160
Cys Pro Ala Gly Trp Glu Gly Ala Thr Cys Asn Ile Ala Arg Asn Ser	
705	
710	
715	
720	
AGC TGC CTG CCA AAC CCC TGT CAC AAT GGT GGT ACC TGT GTA GTT AGT	2208
Ser Cys Leu Pro Asn Pro Cys His Asn Gly Gly Thr Cys Val Val Ser	
725	
730	
735	
GGG GAT TCT TTC ACT TGT GTC TGC AAG GAG GGC TGG GAA GGA CCG ACA	2256
Gly Asp Ser Phe Thr Cys Val Cys Lys Glu Gly Trp Glu Gly Pro Thr	
740	
745	
750	
TGT ACT CAG AAC ACA AAT GAC TGC AGT CCT CAT CCT TGT TAC AAC AGT	2304
Cys Thr Gln Asn Thr Asn Asp Cys Ser Pro His Pro Cys Tyr Asn Ser	
755	
760	
765	
GGT ACT TGT GTG GAT GGA GAC AAC TGG TAC CGC TGT GAG TGC GCT CCC	2352
Gly Thr Cys Val Asp Gly Asp Asn Trp Tyr Arg Cys Glu Cys Ala Pro	
770	
775	
780	
GGC TTC GCA GGT CCC GAC TGT AGG ATC AAC ATC AAT GAA TGT CAG TCT	2400
Gly Phe Ala Gly Pro Asp Cys Arg Ile Asn Ile Asn Glu Cys Gln Ser	
785	
790	
795	
800	
TCA CCC TGT GCC TTT GGG GCT ACT TGT GTG GAT GAA ATT AAT GGG TAC	2448
Ser Pro Cys Ala Phe Gly Ala Thr Cys Val Asp Glu Ile Asn Gly Tyr	
805	
810	
815	
CGT TGC ATT TGT CCA CCG GGT CGC AGT GGT CCA GGA TGC CAG GAA GTT	2496
Arg Cys Ile Cys Pro Pro Gly Arg Ser Gly Pro Gly Cys Gln Glu Val	
820	
825	
830	
ACA GGG AGG CCT TGC TTT ACC AGT ATT CGA GTA ATG CCA GAC GGT GCT	2544
Thr Gly Arg Pro Cys Phe Thr Ser Ile Arg Val Met Pro Asp Gly Ala	
835	
840	
845	

AAG TGG GAT GAT GAC TGT AAT ACT TGT CAG TGT TTG AAT GGA AAA GTC Lys Trp Asp Asp Asp Cys Asn Thr Cys Gln Cys Leu Asn Gly Lys Val 850 855 860	2592
ACC TGT TCT AAG GTT TGG TGT GGT CCT CGA CCT TGT ATA ATA CAT GCC Thr Cys Ser Lys Val Trp Cys Gly Pro Arg Pro Cys Ile Ile His Ala 865 870 875 880	2640
AAA GGT CAT AAT GAA TGC CCA GCT GGA CAC GCT TGT GTT CCT GTT AAA Lys Gly His Asn Glu Cys Pro Ala Gly His Ala Cys Val Pro Val Lys 885 890 895	2688
GAA GAC CAT TGT TTC ACT CAT CCT TGT GCT GCA GTG GGT GAA TGC TGG Glu Asp His Cys Phe Thr His Pro Cys Ala Ala Val Gly Glu Cys Trp 900 905 910	2736
CCT TCT AAT CAG CAG CCT GTG AAG ACC AAA TGC AAT TCT GAT TCT TAT Pro Ser Asn Gln Gln Pro Val Lys Thr Lys Cys Asn Ser Asp Ser Tyr 915 920 925	2784
TAC CAA GAT AAT TGT GCC AAC ATC ACC TTC ACC TTT AAT AAG GAA ATG Tyr Gln Asp Asn Cys Ala Asn Ile Thr Phe Thr Phe Asn Lys Glu Met 930 935 940	2832
ATG GCA CCA GGC CTT ACC ACG GAG CAC ATT TGC AGT GAA TTG AGG AAT Met Ala Pro Gly Leu Thr Thr Glu His Ile Cys Ser Glu Leu Arg Asn 945 950 955 960	2880
CTG AAT ATC CTG AAG AAT GTT TCT GCT GAA TAT TCC ATC TAT ATT ACC Leu Asn Ile Leu Lys Asn Val Ser Ala Glu Tyr Ser Ile Tyr Ile Thr 965 970 975	2928
TGT GAG CCT TCA CAC TTG GCA AAT AAT GAA ATA CAT GTT GCT ATT TCT Cys Glu Pro Ser His Leu Ala Asn Asn Glu Ile His Val Ala Ile Ser 980 985 990	2976
GCT GAA GAT ATA GGA GAA GAT GAA AAC CCA ATC AAG GAA ATC ACA GAT Ala Glu Asp Ile Gly Glu Asp Glu Asn Pro Ile Lys Glu Ile Thr Asp 995 1000 1005	3024
AAG ATT ATT GAC CTT GTC AGT AAG CGT GAT GGA AAC AAC ACA CTA ATT Lys Ile Ile Asp Leu Val Ser Lys Arg Asp Gly Asn Asn Thr Leu Ile 1010 1015 1020	3072
GCT GCA GTC GCA GAA GTC AGA GTA CAA AGG CGA CCA GTT AAG AAC AAA Ala Ala Val Ala Glu Val Arg Val Gln Arg Pro Val Lys Asn Lys 1025 1030 1035 1040	3120
ACA GAT TTC TTG GTG CCA TTA CTG AGC TCA GTC TTA ACA GTA GCC TGG Thr Asp Phe Leu Val Pro Leu Leu Ser Ser Val Leu Thr Val Ala Trp 1045 1050 1055	3168
ATC TGC TGT CTG GTA ACT GTT TTC TAT TGG TGC ATT CAA AAG CGC AGA Ile Cys Cys Leu Val Thr Val Phe Tyr Trp Cys Ile Gln Lys Arg Arg 1060 1065 1070	3216
AAG CAG AGC AGC CAT ACT CAC ACA GCA TCT GAT GAC AAC ACC ACC AAC Lys Gln Ser Ser His Thr His Thr Ala Ser Asp Asp Asn Thr Thr Asn 1075 1080 1085	3264
AAC GTA AGG GAG CAG CTG AAT CAG ATT AAA AAC CCC ATA GAG AAA CAC Asn Val Arg Glu Gln Leu Asn Gln Ile Lys Asn Pro Ile Glu Lys His 1090 1095 1100	3312
GGA GCA AAT ACT GTT CCA ATT AAA GAC TAT GAA AAC AAA AAC TCT AAA Gly Ala Asn Thr Val Pro Ile Lys Asp Tyr Glu Asn Lys Asn Ser Lys 1105 1110 1115 1120	3360

ATC GCC AAA ATA AGG ACG CAC AAT TCA GAA GTG GAG GAA GAT GAC ATG 3408  
 Ile Ala Lys Ile Arg Thr His Asn Ser Glu Val Glu Glu Asp Asp Met  
                   1125                  1130                  1135

GAC AAA CAC CAG CAA AAG GCC CGG TTT GCC AAG CAG CCA GCG TAC ACT 3456  
 Asp Lys His Gln Gln Lys Ala Arg Phe Ala Lys Gln Pro Ala Tyr Thr  
                   1140                  1145                  1150

TTG GTA GAC AGA GAT GAA AAG CCA CCC AAC AGC ACA CCC ACA AAA CAC 3504  
 Leu Val Asp Arg Asp Glu Lys Pro Pro Asn Ser Thr Pro Thr Lys His  
                   1155                  1160                  1165

CCA AAC TGG ACA AAT AAA CAG GAC AAC AGA GAC TTG GAA AGT GCA CAA 3552  
 Pro Asn Trp Thr Asn Lys Gln Asp Asn Arg Asp Leu Glu Ser Ala Gln  
                   1170                  1175                  1180

AGT TTA AAT AGA ATG GAG TAC ATT GTA  
 Ser Leu Asn Arg Met Glu Tyr Ile Val  
 1185                  1190

## (2) INFORMATION FOR SEQ ID NO:6:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1194 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Gln Val Ala Ser Ala Ser Gly Gln Phe Glu Leu Glu Ile Leu Ser Val  
 1                  5                  10                  15

Gln Asn Val Asn Gly Val Leu Gln Asn Gly Asn Cys Cys Asp Gly Thr  
                   20                  25                  30

Arg Asn Pro Gly Asp Lys Lys Cys Thr Arg Asp Glu Cys Asp Thr Tyr  
                   35                  40                  45

Phe Lys Val Cys Leu Lys Glu Tyr Gln Ser Arg Val Thr Ala Gly Gly  
                   50                  55                  60

Pro Cys Ser Phe Gly Ser Lys Ser Thr Pro Val Ile Gly Gly Asn Thr  
                   65                  70                  75                  80

Phe Asn Leu Lys Tyr Ser Arg Asn Asn Glu Lys Asn Arg Ile Val Ile  
                   85                  90                  95

Pro Phe Thr Phe Ala Trp Pro Arg Ser Tyr Thr Leu Leu Val Glu Ala  
                   100                  105                  110

Trp Asp Tyr Asn Asp Asn Ser Thr Asn Pro Asp Arg Ile Ile Glu Lys  
                   115                  120                  125

Ala Ser His Ser Gly Met Ile Asn Pro Ser Arg Gln Trp Gln Thr Leu  
                   130                  135                  140

Lys His Asn Thr Gly Ala Ala His Phe Glu Tyr Gln Ile Arg Val Thr  
                   145                  150                  155                  160

Cys Ala Glu His Tyr Tyr Gly Phe Gly Cys Asn Lys Phe Cys Arg Pro  
                   165                  170                  175

Arg Asp Asp Phe Phe Thr His His Thr Cys Asp Gln Asn Gly Asn Lys  
 180 185 190  
 Thr Cys Leu Glu Gly Trp Thr Gly Pro Glu Cys Asn Lys Ala Ile Cys  
 195 200 205  
 Arg Gln Gly Cys Ser Pro Lys His Gly Ser Cys Thr Val Pro Gly Glu  
 210 215 220  
 Cys Arg Cys Gln Tyr Gly Trp Gln Gly Gln Tyr Cys Asp Lys Cys Ile  
 225 230 235 240  
 Pro His Pro Gly Cys Val His Gly Thr Cys Ile Glu Pro Trp Gln Cys  
 245 250 255  
 Leu Cys Glu Thr Asn Trp Gly Gly Gln Leu Cys Asp Lys Asp Leu Asn  
 260 265 270  
 Tyr Cys Gly Thr His Pro Pro Cys Leu Asn Gly Gly Thr Cys Ser Asn  
 275 280 285  
 Thr Gly Pro Asp Lys Tyr Gln Cys Ser Cys Pro Glu Gly Tyr Ser Gly  
 290 295 300  
 Gln Asn Cys Glu Ile Ala Glu His Ala Cys Leu Ser Asp Pro Cys His  
 305 310 315 320  
 Asn Gly Gly Ser Cys Leu Glu Thr Ser Thr Gly Phe Glu Cys Val Cys  
 325 330 335  
 Ala Pro Gly Trp Ala Gly Pro Thr Cys Thr Asp Asn Ile Asp Asp Cys  
 340 345 350  
 Ser Pro Asn Pro Cys Gly His Gly Gly Thr Cys Gln Asp Leu Val Asp  
 355 360 365  
 Gly Phe Lys Cys Ile Cys Pro Pro Gln Trp Thr Gly Lys Thr Cys Gln  
 370 375 380  
 Leu Asp Ala Asn Glu Cys Glu Gly Lys Pro Cys Val Asn Ala Asn Ser  
 385 390 395 400  
 Cys Arg Asn Leu Ile Gly Ser Tyr Tyr Cys Asp Cys Ile Thr Gly Trp  
 405 410 415  
 Ser Gly His Asn Cys Asp Ile Asn Ile Asn Asp Cys Arg Gly Gln Cys  
 420 425 430  
 Gln Asn Gly Gly Ser Cys Arg Asp Leu Val Asn Gly Tyr Arg Cys Ile  
 435 440 445  
 Cys Ser Pro Gly Tyr Ala Gly Asp His Cys Glu Lys Asp Ile Asn Glu  
 450 455 460  
 Cys Ala Ser Asn Pro Cys Met Asn Gly Gly His Cys Gln Asp Glu Ile  
 465 470 475 480  
 Asn Gly Phe Gln Cys Leu Cys Pro Ala Gly Phe Ser Gly Asn Leu Cys  
 485 490 495  
 Gln Leu Asp Ile Asp Tyr Cys Glu Pro Asn Pro Cys Gln Asn Gly Ala  
 500 505 510  
 Gln Cys Phe Asn Leu Ala Met Asp Tyr Phe Cys Asn Cys Pro Glu Asp  
 515 520 525  
 Tyr Glu Gly Lys Asn Cys Ser His Leu Lys Asp His Cys Arg Thr Thr



530					535					540					
Pro	Cys	Glu	Val	Ile	Asp	Ser	Cys	Thr	Val	Ala	Val	Ala	Ser	Asn	Ser
545					550					555					560
Thr	Pro	Glu	Gly	Val	Arg	Tyr	Ile	Ser	Ser	Asn	Val	Cys	Gly	Pro	His
				565					570					575	
Gly	Lys	Cys	Lys	Ser	Gln	Ala	Gly	Gly	Lys	Phe	Thr	Cys	Glu	Cys	Asn
			580					585					590		
Lys	Gly	Phe	Thr	Gly	Thr	Tyr	Cys	His	Glu	Asn	Ile	Asn	Asp	Cys	Glu
		595					600					605			
Ser	Asn	Pro	Cys	Lys	Asn	Gly	Gly	Thr	Cys	Ile	Asp	Gly	Val	Asn	Ser
	610					615					620				
Tyr	Lys	Cys	Ile	Cys	Ser	Asp	Gly	Trp	Glu	Gly	Thr	Tyr	Cys	Glu	Thr
625					630					635					640
Asn	Ile	Asn	Asp	Cys	Ser	Lys	Asn	Pro	Cys	His	Asn	Gly	Gly	Thr	Cys
				645					650					655	
Arg	Asp	Leu	Val	Asn	Asp	Phe	Phe	Cys	Glu	Cys	Lys	Asn	Gly	Trp	Lys
			660					665					670		
Gly	Lys	Thr	Cys	His	Ser	Arg	Asp	Ser	Gln	Cys	Asp	Glu	Ala	Thr	Cys
		675					680					685			
Asn	Asn	Gly	Gly	Thr	Cys	Tyr	Asp	Glu	Gly	Asp	Thr	Phe	Lys	Cys	Met
	690					695					700				
Cys	Pro	Ala	Gly	Trp	Glu	Gly	Ala	Thr	Cys	Asn	Ile	Ala	Arg	Asn	Ser
705					710					715					720
Ser	Cys	Leu	Pro	Asn	Pro	Cys	His	Asn	Gly	Gly	Thr	Cys	Val	Val	Ser
				725					730					735	
Gly	Asp	Ser	Phe	Thr	Cys	Val	Cys	Lys	Glu	Gly	Trp	Glu	Gly	Pro	Thr
			740					745					750		
Cys	Thr	Gln	Asn	Thr	Asn	Asp	Cys	Ser	Pro	His	Pro	Cys	Tyr	Asn	Ser
		755					760					765			
Gly	Thr	Cys	Val	Asp	Gly	Asp	Asn	Trp	Tyr	Arg	Cys	Glu	Cys	Ala	Pro
	770					775					780				
Gly	Phe	Ala	Gly	Pro	Asp	Cys	Arg	Ile	Asn	Ile	Asn	Glu	Cys	Gln	Ser
785						790					795				800
Ser	Pro	Cys	Ala	Phe	Gly	Ala	Thr	Cys	Val	Asp	Glu	Ile	Asn	Gly	Tyr
				805					810					815	
Arg	Cys	Ile	Cys	Pro	Pro	Gly	Arg	Ser	Gly	Pro	Gly	Cys	Gln	Glu	Val
			820					825					830		
Thr	Gly	Arg	Pro	Cys	Phe	Thr	Ser	Ile	Arg	Val	Met	Pro	Asp	Gly	Ala
		835					840					845			
Lys	Trp	Asp	Asp	Asp	Cys	Asn	Thr	Cys	Gln	Cys	Leu	Asn	Gly	Lys	Val
	850					855					860				
Thr	Cys	Ser	Lys	Val	Trp	Cys	Gly	Pro	Arg	Pro	Cys	Ile	Ile	His	Ala
865						870					875				880
Lys	Gly	His	Asn	Glu	Cys	Pro	Ala	Gly	His	Ala	Cys	Val	Pro	Val	Lys
				885					890					895	

Glu Asp His Cys Phe Thr His Pro Cys Ala Ala Val Gly Glu Cys Trp  
                   900                                  905                                  910  
 Pro Ser Asn Gln Gln Pro Val Lys Thr Lys Cys Asn Ser Asp Ser Tyr  
                   915                                  920                                  925  
 Tyr Gln Asp Asn Cys Ala Asn Ile Thr Phe Thr Phe Asn Lys Glu Met  
                   930                                  935                                  940  
 Met Ala Pro Gly Leu Thr Thr Glu His Ile Cys Ser Glu Leu Arg Asn  
                   945                                  950                                  955                                  960  
 Leu Asn Ile Leu Lys Asn Val Ser Ala Glu Tyr Ser Ile Tyr Ile Thr  
                   965                                  970                                  975  
 Cys Glu Pro Ser His Leu Ala Asn Asn Glu Ile His Val Ala Ile Ser  
                   980                                  985                                  990  
 Ala Glu Asp Ile Gly Glu Asp Glu Asn Pro Ile Lys Glu Ile Thr Asp  
                   995                                  1000                                  1005  
 Lys Ile Ile Asp Leu Val Ser Lys Arg Asp Gly Asn Asn Thr Leu Ile  
                   1010                                  1015                                  1020  
 Ala Ala Val Ala Glu Val Arg Val Gln Arg Arg Pro Val Lys Asn Lys  
                   1025                                  1030                                  1035                                  1040  
 Thr Asp Phe Leu Val Pro Leu Leu Ser Ser Val Leu Thr Val Ala Trp  
                   1045                                  1050                                  1055  
 Ile Cys Cys Leu Val Thr Val Phe Tyr Trp Cys Ile Gln Lys Arg Arg  
                   1060                                  1065                                  1070  
 Lys Gln Ser Ser His Thr His Thr Ala Ser Asp Asp Asn Thr Thr Asn  
                   1075                                  1080                                  1085  
 Asn Val Arg Glu Gln Leu Asn Gln Ile Lys Asn Pro Ile Glu Lys His  
                   1090                                  1095                                  1100  
 Gly Ala Asn Thr Val Pro Ile Lys Asp Tyr Glu Asn Lys Asn Ser Lys  
                   1105                                  1110                                  1115                                  1120  
 Ile Ala Lys Ile Arg Thr His Asn Ser Glu Val Glu Glu Asp Asp Met  
                   1125                                  1130                                  1135  
 Asp Lys His Gln Gln Lys Ala Arg Phe Ala Lys Gln Pro Ala Tyr Thr  
                   1140                                  1145                                  1150  
 Leu Val Asp Arg Asp Glu Lys Pro Pro Asn Ser Thr Pro Thr Lys His  
                   1155                                  1160                                  1165  
 Pro Asn Trp Thr Asn Lys Gln Asp Asn Arg Asp Leu Glu Ser Ala Gln  
                   1170                                  1175                                  1180  
 Ser Leu Asn Arg Met Glu Tyr Ile Val  
                   1185                                  1190

## (2) INFORMATION FOR SEQ ID NO:7:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 236 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Met His Trp Ile Lys Cys Leu Leu Thr Ala Phe Ile Cys Phe Thr Val  
 1 5 10 15  
 Ile Val Gln Val His Ser Ser Gly Ser Phe Glu Leu Arg Leu Lys Tyr  
 20 25 30  
 Phe Ser Asn Asp His Gly Arg Asp Asn Glu Gly Arg Cys Cys Ser Gly  
 35 40 45  
 Glu Ser Asp Gly Ala Thr Gly Lys Cys Leu Gly Ser Cys Lys Thr Arg  
 50 55 60  
 Phe Arg Val Cys Leu Lys His Tyr Gln Ala Thr Ile Asp Thr Thr Ser  
 65 70 75 80  
 Gln Cys Thr Tyr Gly Asp Val Ile Thr Pro Ile Leu Gly Glu Asn Ser  
 85 90 95  
 Val Asn Leu Thr Asp Ala Gln Arg Phe Gln Asn Lys Gly Phe Thr Asn  
 100 105 110  
 Pro Ile Gln Phe Pro Phe Ser Phe Ser Trp Pro Gly Thr Phe Ser Leu  
 115 120 125  
 Ile Val Glu Ala Trp His Asp Thr Asn Asn Ser Gly Asn Ala Arg Thr  
 130 135 140  
 Asn Lys Leu Leu Ile Gln Arg Leu Leu Val Gln Gln Val Leu Glu Val  
 145 150 155 160  
 Ser Ser Glu Trp Lys Thr Asn Lys Ser Glu Ser Gln Tyr Thr Ser Leu  
 165 170 175  
 Glu Tyr Asp Phe Arg Val Thr Cys Asp Leu Asn Tyr Tyr Gly Ser Gly  
 180 185 190  
 Cys Ala Lys Phe Cys Arg Pro Arg Asp Asp Ser Phe Gly His Ser Thr  
 195 200 205  
 Cys Ser Glu Thr Gly Glu Ile Ile Cys Leu Thr Gly Trp Gln Gly Asp  
 210 215 220  
 Tyr Cys His Ile Pro Lys Cys Ala Lys Gly Cys Glu  
 225 230 235

## (2) INFORMATION FOR SEQ ID NO:8:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1405 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met Phe Arg Lys His Phe Arg Arg Lys Pro Ala Thr Ser Ser Ser Leu  
 1 5 10 15  
 Glu Ser Thr Ile Glu Ser Ala Asp Ser Leu Gly Met Ser Lys Lys Thr  
 20 25 30  
 Ala Thr Lys Arg Gln Arg Pro Arg His Arg Val Pro Lys Ile Ala Thr  
 35 40 45  
 Leu Pro Ser Thr Ile Arg Asp Cys Arg Ser Leu Lys Ser Ala Cys Asn  
 50 55 60

Leu Ile Ala Leu Ile Leu Ile Leu Leu Val His Lys Ile Ser Ala Ala  
 65 70 75 80  
 Gly Asn Phe Glu Leu Glu Ile Leu Glu Ile Ser Asn Thr Asn Ser His  
 85 90 95  
 Leu Leu Asn Gly Tyr Cys Cys Gly Met Pro Ala Glu Leu Arg Ala Thr  
 100 105 110  
 Lys Thr Ile Gly Cys Ser Pro Cys Thr Thr Ala Phe Arg Leu Cys Leu  
 115 120 125  
 Lys Glu Tyr Gln Thr Thr Glu Gln Gly Ala Ser Ile Ser Thr Gly Cys  
 130 135 140  
 Ser Phe Gly Asn Ala Thr Thr Lys Ile Leu Gly Gly Ser Ser Phe Val  
 145 150 155 160  
 Leu Ser Asp Pro Gly Val Gly Ala Ile Val Leu Pro Phe Thr Phe Arg  
 165 170 175  
 Trp Thr Lys Ser Phe Thr Leu Ile Leu Gln Ala Leu Asp Met Tyr Asn  
 180 185 190  
 Thr Ser Tyr Pro Asp Ala Glu Arg Leu Ile Glu Glu Thr Ser Tyr Ser  
 195 200 205  
 Gly Val Ile Leu Pro Ser Pro Glu Trp Lys Thr Leu Asp His Ile Gly  
 210 215 220  
 Arg Asn Ala Arg Ile Thr Tyr Arg Val Arg Val Gln Cys Ala Val Thr  
 225 230 235 240  
 Tyr Tyr Asn Thr Thr Cys Thr Thr Phe Cys Arg Pro Arg Asp Asp Gln  
 245 250 255  
 Phe Gly His Tyr Ala Cys Gly Ser Glu Gly Gln Lys Leu Cys Leu Asn  
 260 265 270  
 Gly Trp Gln Gly Val Asn Cys Glu Glu Ala Ile Cys Lys Ala Gly Cys  
 275 280 285  
 Asp Pro Val His Gly Lys Cys Asp Arg Pro Gly Glu Cys Glu Cys Arg  
 290 295 300  
 Pro Gly Trp Arg Gly Pro Leu Cys Asn Glu Cys Met Val Tyr Pro Gly  
 305 310 315 320  
 Cys Lys His Gly Ser Cys Asn Gly Ser Ala Trp Lys Cys Val Cys Asp  
 325 330 335  
 Thr Asn Trp Gly Gly Ile Leu Cys Asp Gln Asp Leu Asn Phe Cys Gly  
 340 345 350  
 Thr His Glu Pro Cys Lys His Gly Gly Thr Cys Glu Asn Thr Ala Pro  
 355 360 365  
 Asp Lys Tyr Arg Cys Thr Cys Ala Glu Gly Leu Ser Gly Glu Gln Cys  
 370 375 380  
 Glu Ile Val Glu His Pro Cys Ala Thr Arg Pro Cys Arg Asn Gly Gly  
 385 390 395 400  
 Thr Cys Thr Leu Lys Thr Ser Asn Arg Thr Gln Ala Gln Val Tyr Arg  
 405 410 415  
 Thr Ser His Gly Arg Ser Asn Met Gly Arg Pro Val Arg Arg Ser Ser

420										425					430				
Ser	Met	Arg	Ser	Leu	Asp	His	Leu	Arg	Pro	Glu	Gly	Gln	Ala	Leu	Asn				
		435						440					445						
Gly	Ser	Ser	Ser	Ser	Gly	Leu	Val	Ser	Leu	Gly	Ser	Leu	Gln	Leu	Gln				
	450					455						460							
Gln	Gln	Leu	Ala	Pro	Asp	Phe	Thr	Cys	Asp	Cys	Ala	Ala	Gly	Trp	Thr				
465					470					475					480				
Gly	Pro	Thr	Cys	Glu	Ile	Asn	Ile	Asp	Glu	Cys	Ala	Gly	Gly	Pro	Cys				
				485					490					495					
Glu	His	Gly	Gly	Thr	Cys	Ile	Asp	Leu	Ile	Gly	Gly	Phe	Arg	Cys	Glu				
			500					505					510						
Cys	Pro	Pro	Glu	Trp	His	Gly	Asp	Val	Cys	Gln	Val	Asp	Val	Asn	Glu				
		515					520					525							
Cys	Glu	Ala	Pro	His	Ser	Ala	Gly	Ile	Ala	Ala	Asn	Ala	Leu	Leu	Thr				
	530					535					540								
Thr	Thr	Ala	Thr	Ala	Ile	Ile	Gly	Ser	Asn	Leu	Ser	Ser	Thr	Ala	Leu				
545					550					555					560				
Leu	Ala	Ala	Leu	Thr	Ser	Ala	Val	Ala	Ser	Thr	Ser	Leu	Ala	Ile	Gly				
				565					570					575					
Pro	Cys	Ile	Asn	Ala	Lys	Glu	Cys	Arg	Asn	Gln	Pro	Gly	Ser	Phe	Ala				
			580					585					590						
Cys	Ile	Cys	Lys	Glu	Gly	Trp	Gly	Gly	Val	Thr	Cys	Ala	Glu	Asn	Leu				
	595						600					605							
Asp	Asp	Cys	Val	Gly	Gln	Cys	Arg	Asn	Gly	Ala	Thr	Cys	Ile	Asp	Leu				
	610					615				620									
Val	Asn	Asp	Tyr	Arg	Cys	Ala	Cys	Ala	Ser	Gly	Phe	Thr	Gly	Arg	Asp				
625					630					635					640				
Cys	Glu	Thr	Asp	Ile	Asp	Glu	Cys	Ala	Thr	Ser	Pro	Cys	Arg	Asn	Gly				
				645					650					655					
Gly	Glu	Cys	Val	Asp	Met	Val	Gly	Lys	Phe	Asn	Cys	Ile	Cys	Pro	Leu				
			660					665					670						
Gly	Tyr	Ser	Gly	Ser	Leu	Cys	Glu	Glu	Ala	Lys	Glu	Asn	Cys	Thr	Pro				
		675					680						685						
Ser	Pro	Cys	Leu	Glu	Gly	His	Cys	Leu	Asn	Thr	Pro	Glu	Gly	Tyr	Tyr				
	690					695					700								
Cys	His	Cys	Pro	Pro	Asp	Arg	Ala	Gly	Lys	His	Cys	Glu	Gln	Leu	Arg				
705					710					715					720				
Pro	Leu	Cys	Ser	Gln	Pro	Pro	Cys	Asn	Glu	Gly	Cys	Phe	Ala	Asn	Val				
				725					730					735					
Ser	Leu	Ala	Thr	Ser	Ala	Thr	Thr	Thr	Thr	Thr	Thr	Thr	Thr	Thr	Ala				
			740					745						750					
Thr	Thr	Thr	Arg	Lys	Met	Ala	Lys	Pro	Ser	Gly	Leu	Pro	Cys	Ser	Gly				
		755					760					765							
His	Gly	Ser	Cys	Glu	Met	Ser	Asp	Val	Gly	Thr	Phe	Cys	Lys	Cys	His				
	770					775						780							

Val Gly His Thr Gly Thr Phe Cys Glu His Asn Leu Asn Glu Cys Ser  
 785 790 795 800  
 Pro Asn Pro Cys Arg Asn Gly Gly Ile Cys Leu Asp Gly Asp Gly Asp  
 805 810 815  
 Phe Thr Cys Glu Cys Met Ser Gly Trp Thr Gly Lys Arg Cys Ser Glu  
 820 825 830  
 Arg Ala Thr Gly Cys Tyr Ala Gly Gln Cys Gln Asn Gly Gly Thr Cys  
 835 840 845  
 Met Pro Gly Ala Pro Asp Lys Ala Leu Gln Pro His Cys Arg Cys Ala  
 850 855 860  
 Pro Gly Trp Thr Gly Leu Phe Cys Ala Glu Ala Ile Asp Gln Cys Arg  
 865 870 875 880  
 Gly Gln Pro Cys His Asn Gly Gly Thr Cys Glu Ser Gly Ala Gly Trp  
 885 890 895  
 Phe Arg Cys Val Cys Ala Gln Gly Phe Ser Gly Pro Asp Cys Arg Ile  
 900 905 910  
 Asn Val Asn Glu Cys Ser Pro Gln Pro Cys Gln Gly Gly Ala Thr Cys  
 915 920 925  
 Ile Asp Gly Ile Gly Gly Tyr Ser Cys Ile Cys Pro Pro Gly Arg His  
 930 935 940  
 Gly Leu Arg Cys Glu Ile Leu Leu Ser Asp Pro Lys Ser Ala Cys Gln  
 945 950 955 960  
 Asn Ala Ser Asn Thr Ile Ser Pro Tyr Thr Ala Leu Asn Arg Ser Gln  
 965 970 975  
 Asn Trp Leu Asp Ile Ala Leu Thr Gly Arg Thr Glu Asp Asp Glu Asn  
 980 985 990  
 Cys Asn Ala Cys Val Cys Glu Asn Gly Thr Ser Arg Cys Thr Asn Leu  
 995 1000 1005  
 Trp Cys Gly Leu Pro Asn Cys Tyr Lys Val Asp Pro Leu Ser Lys Ser  
 1010 1015 1020  
 Ser Asn Leu Ser Gly Val Cys Lys Gln His Glu Val Cys Val Pro Ala  
 1025 1030 1035 1040  
 Leu Ser Glu Thr Cys Leu Ser Ser Pro Cys Asn Val Arg Gly Asp Cys  
 1045 1050 1055  
 Arg Ala Leu Glu Pro Ser Arg Arg Val Ala Pro Pro Arg Leu Pro Ala  
 1060 1065 1070  
 Lys Ser Ser Cys Trp Pro Asn Gln Ala Val Val Asn Glu Asn Cys Ala  
 1075 1080 1085  
 Arg Leu Thr Ile Leu Leu Ala Leu Glu Arg Val Gly Lys Gly Ala Ser  
 1090 1095 1100  
 Val Glu Gly Leu Cys Ser Leu Val Arg Val Leu Leu Ala Ala Gln Leu  
 1105 1110 1115 1120  
 Ile Lys Lys Pro Ala Ser Thr Phe Gly Gln Asp Pro Gly Met Leu Met  
 1125 1130 1135  
 Val Leu Cys Asp Leu Lys Thr Gly Thr Asn Asp Thr Val Glu Leu Thr

1140 1145 1150  
 Val Ser Ser Ser Lys Leu Asn Asp Pro Gln Leu Pro Val Ala Val Gly  
 1155 1160 1165  
 Leu Leu Gly Glu Leu Leu Ser Ser Arg Gln Leu Asn Gly Ile Gln Arg  
 1170 1175 1180  
 Arg Lys Glu Leu Glu Leu Gln His Ala Lys Leu Ala Ala Leu Thr Ser  
 1185 1190 1195 1200  
 Ile Val Glu Val Lys Leu Glu Thr Ala Arg Val Ala Asp Gly Ser Gly  
 1205 1210 1215  
 His Ser Leu Leu Ile Gly Val Leu Cys Gly Val Phe Ile Val Leu Val  
 1220 1225 1230  
 Gly Phe Ser Val Phe Ile Ser Leu Tyr Trp Lys Gln Arg Leu Ala Tyr  
 1235 1240 1245  
 Arg Thr Ser Ser Gly Met Asn Leu Thr Pro Ser Leu Asp Ala Leu Arg  
 1250 1255 1260  
 His Glu Glu Glu Lys Ser Asn Asn Leu Gln Asn Glu Glu Asn Leu Arg  
 1265 1270 1275 1280  
 Arg Tyr Thr Asn Pro Leu Lys Gly Ser Thr Ser Ser Leu Arg Ala Ala  
 1285 1290 1295  
 Thr Gly Met Glu Leu Ser Leu Asn Pro Ala Pro Glu Leu Ala Ala Ser  
 1300 1305 1310  
 Ala Ala Ser Ser Ser Ala Leu His Arg Ser Gln Pro Leu Phe Pro Pro  
 1315 1320 1325  
 Cys Asp Phe Glu Arg Glu Leu Asp Ser Ser Thr Gly Leu Lys Gln Ala  
 1330 1335 1340  
 His Lys Arg Ser Ser Gln Ile Leu Leu His Lys Thr Gln Asn Ser Asp  
 1345 1350 1355 1360  
 Met Arg Lys Asn Thr Val Gly Ser Leu Asp Ser Pro Arg Lys Asp Phe  
 1365 1370 1375  
 Gly Lys Arg Ser Ile Asn Cys Lys Ser Met Pro Pro Ser Ser Gly Asp  
 1380 1385 1390  
 Glu Gly Ser Asp Val Leu Ala Thr Thr Val Met Val  
 1395 1400

## (2) INFORMATION FOR SEQ ID NO:9:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA

## (ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 3
- (D) OTHER INFORMATION: /mod\_base= i

## (ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 12
- (D) OTHER INFORMATION: /mod\_base= i

(ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 18
- (D) OTHER INFORMATION: /mod\_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CGNYTTTGCY TNAARSANTA YCA

23

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
  - (B) LOCATION: 6
  - (D) OTHER INFORMATION: /label= A
- /note= "X=histidine or glutamic acid"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Arg Leu Cys Cys Lys Xaa Tyr Gln  
1 5

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 3
- (D) OTHER INFORMATION: /mod\_base= i

(ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 9
- (D) OTHER INFORMATION: /mod\_base= i

(ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 12
- (D) OTHER INFORMATION: /mod\_base= i

(ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 15
- (D) OTHER INFORMATION: /mod\_base= i



(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

TCNATGCANG TNCCNCCRTT

20

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 7 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Asn Gly Gly Thr Cys Ile Asp  
 1 5

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 163 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

- (A) NAME/KEY: CDS  
 (B) LOCATION: 2..163

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

G TCC CGC GTC ACT GCC GGG GGA CCC TGC AGC TTC GGC TCA GGG TCT	46
Ser Arg Val Thr Ala Gly Gly Pro Cys Ser Phe Gly Ser Gly Ser	
1 5 10 15	
ACG CCT GTC ATC GGG GGT AAC ACC TTC AAT CTC AAG GCC AGC CGT GGC	94
Thr Pro Val Ile Gly Gly Asn Thr Phe Asn Leu Lys Ala Ser Arg Gly	
20 25 30	
AAC GAC CGT AAT CGC ATC GTA CTG CCT TTC AGT TTC ACC TGG CCG AGG	142
Asn Asp Arg Asn Arg Ile Val Leu Pro Phe Ser Phe Thr Trp Pro Arg	
35 40 45	
TCC TAC ACT TTG CTG GTG GAG	163
Ser Tyr Thr Leu Leu Val Glu	
50	

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 54 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Ser Arg Val Thr Ala Gly Gly Pro Cys Ser Phe Gly Ser Gly Ser Thr  
 1 5 10 15  
 Pro Val Ile Gly Gly Asn Thr Phe Asn Leu Lys Ala Ser Arg Gly Asn  
 20 25 30  
 Asp Arg Asn Arg Ile Val Leu Pro Phe Ser Phe Thr Trp Pro Arg Ser  
 35 40 45  
 Tyr Thr Leu Leu Val Glu  
 50

## (2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 135 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

## (ix) FEATURE:

- (A) NAME/KEY: CDS  
 (B) LOCATION: 1..135

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

TCT TCT AAC GTC TGT GGT CCC CAT GGC AAG TGC AAG AGC CAG TCG GCA 48  
 Ser Ser Asn Val Cys Gly Pro His Gly Lys Cys Lys Ser Gln Ser Ala  
 1 5 10 15  
 GGC AAA TTC ACC TGT GAC TGT AAC AAA GGC TTC ACC GGC ACC TAC TGC 96  
 Gly Lys Phe Thr Cys Asp Cys Asn Lys Gly Phe Thr Gly Thr Tyr Cys  
 20 25 30  
 CAT GAA AAT ATC AAC GAC TGC GAG AGC AAC CCC TGT AAA 135  
 His Glu Asn Ile Asn Asp Cys Glu Ser Asn Pro Cys Lys  
 35 40 45

## (2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 45 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Ser Ser Asn Val Cys Gly Pro His Gly Lys Cys Lys Ser Gln Ser Ala  
 1 5 10 15  
 Gly Lys Phe Thr Cys Asp Cys Asn Lys Gly Phe Thr Gly Thr Tyr Cys  
 20 25 30  
 His Glu Asn Ile Asn Asp Cys Glu Ser Asn Pro Cys Lys  
 35 40 45

## (2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 23 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- (ix) FEATURE:
  - (A) NAME/KEY: modified\_base
  - (B) LOCATION: 3
  - (D) OTHER INFORMATION: /mod\_base= i

- (ix) FEATURE:
  - (A) NAME/KEY: modified\_base
  - (B) LOCATION: 6
  - (D) OTHER INFORMATION: /mod\_base= i

- (ix) FEATURE:
  - (A) NAME/KEY: modified\_base
  - (B) LOCATION: 12
  - (D) OTHER INFORMATION: /mod\_base= i

- (ix) FEATURE:
  - (A) NAME/KEY: modified\_base
  - (B) LOCATION: 18
  - (D) OTHER INFORMATION: /mod\_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

CGNYTNTGCY TNAARSANTA YCA

23

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 8 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS:
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- (ix) FEATURE:
    - (A) NAME/KEY: Modified-site
    - (B) LOCATION: 6
    - (D) OTHER INFORMATION: /label= A
- /note= "X=glutamic acid or histidine"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Arg Leu Cys Leu Lys Xaa Tyr Gln  
1 5

International Application No: PCT/ /

**MICROORGANISMS**Optional Sheet in connection with the microorganism referred to on page 86-87, lines 1-40 of the description \***A. IDENTIFICATION OF DEPOSIT \***

Further deposits are identified on an additional sheet \*

Name of depositary institution \*

American Type Culture Collection

Address of depositary institution (including postal code and country) \*

12301 Parklawn Drive  
Rockville, MD 20852  
USDate of deposit \* February 28, 1995 Accession Number \* 97068**B. ADDITIONAL INDICATIONS \*** (leave blank if not applicable). This information is continued on a separate attached sheet**C. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE \*** (if the indications are for all designated States)**D. SEPARATE FURNISHING OF INDICATIONS \*** (leave blank if not applicable)The indications listed below will be submitted to the International Bureau later \* (Specify the general nature of the indications e.g.,  
"Accession Number of Deposit")E. ☐ This sheet was received with the International application when filed (to be checked by the receiving Office)\_\_\_\_\_  
(Authorized Officer)☐ The date of receipt (from the applicant) by the International Bureau \*

was

\_\_\_\_\_  
(Authorized Officer)

Form PCT/RO/134 (January 1981)

International Application No: PCT/ /

Form PCT/RO/134 (cont.)

American Type Culture Collection

12301 Perkiawn Drive  
Rockville, MD 20852  
US

Accession No.

Date of Deposit

March 5, 1996

March 5, 1996

WHAT IS CLAIMED IS:

1. A purified vertebrate Serrate protein.
- 5        2. The protein of claim 1 which is a human protein.
3. The protein of claim 1 which is a mammalian protein.
- 10       4. The protein of claim 2 which comprises the amino acid sequence substantially as set forth in amino acid numbers 30 - 1218 of SEQ ID NO:2.
- 15       5. The protein of claim 2 which comprises the amino acid sequence substantially as set forth in amino acid numbers 1 - 1257 of SEQ ID NO:4.
6. A purified human protein encoded by a nucleic  
20 acid hybridizable to plasmid SerFL or the Serrate sequence therein as deposited with the ATCC and assigned accession number 68876.
7. The protein of claim 2 which is encoded by  
25 plasmid pBS39 as deposited with the ATCC and assigned accession number 97068.
8. The protein of claim 2 which comprises the Serrate amino acid sequence encoded by plasmid pBS15 as  
30 deposited with the ATCC and assigned accession number \_\_\_\_.
9. The protein of claim 2 which comprises the Serrate amino acid sequence encoded by plasmid pBS3-2 as  
deposited with the ATCC and assigned accession number \_\_\_\_.
- 35

10. A purified fragment of the protein of claim 1, which is able to display one or more functional activities of a Serrate protein.
- 5 11. A purified fragment of the protein of claim 2, which is able to display one or more functional activities of a human or *D. melanogaster* Serrate protein.
12. A purified fragment of the protein of claim 2 10 or 7, which is able to be bound by an antibody directed against a human Serrate protein.
13. A molecule comprising the fragment of claim 10.
- 15 14. A purified fragment of a vertebrate Serrate protein comprising a domain of the protein selected from the group consisting of the extracellular domain, DSL domain, epidermal growth factor-like repeat domain, cysteine-rich 20 domain, transmembrane domain, and intracellular domain.
15. A purified fragment of a vertebrate Serrate protein comprising the DSL domain of the protein.
- 25 16. A purified fragment of a vertebrate Serrate protein comprising an epidermal growth factor-homologous repeat of the protein.
17. The fragment of claim 14 in which the Serrate 30 protein is a human Serrate protein.
18. A purified fragment of a vertebrate Serrate protein comprising a region homologous to a Notch protein or a Delta protein, and consisting of at least ten amino acids.
- 35 19. A chimeric protein comprising a fragment of a vertebrate Serrate protein consisting of at least ten amino

acids fused via a covalent bond to an amino acid sequence of a second protein, in which the second protein is not a Serrate protein.

5           20. The chimeric protein of claim 19 in which the fragment of a Serrate protein is a fragment capable of being bound by an anti-Serrate antibody.

          21. The chimeric protein of claim 19 in which the  
10 Serrate protein is a human protein.

          22. The chimeric protein of claim 19 which is able to display one or more functional activities of a Serrate protein.

15

          23. A purified fragment of a vertebrate Serrate protein which fragment (a) is capable of being bound by an anti-Serrate antibody; (b) lacks the transmembrane and intracellular domains of the protein; and (c) consists of at  
20 least ten amino acids of the Serrate protein.

          24. A purified fragment of a vertebrate Serrate protein which fragment (a) is capable of being bound by an anti-Serrate antibody; (b) lacks the extracellular domain of  
25 the protein; and (c) consists of at least ten amino acids of the Serrate protein.

          25. A purified fragment of a vertebrate Serrate protein which is able to bind to a Notch protein.  
30

          26. The fragment of claim 25, which lacks the epidermal growth factor-like repeats of the Serrate protein.

          27. The fragment of claim 23, 24, 25 or 26 in  
35 which the Serrate protein is a human Serrate protein.



28. The fragment of claim 29, which is a fragment of SEQ ID NO:2 or SEQ ID NO:4.

29. A molecule comprising the fragment of claim 5 25.

30. An antibody which is capable of binding the Serrate protein of claim 1 and which does not bind a *Drosophila* Serrate protein.

10

31. An antibody which is capable of binding the Serrate protein of claim 2 and which does not bind a *Drosophila* Serrate protein.

15 32. The antibody of claim 30 which is monoclonal.

33. A molecule comprising a fragment of the antibody of claim 32, which fragment is capable of binding a vertebrate Serrate protein.

20

34. An isolated nucleic acid comprising a nucleotide sequence encoding a vertebrate Serrate protein.

35. The nucleic acid of claim 34 which is DNA.

25

36. An isolated nucleic acid comprising a nucleotide sequence absolutely complementary to the nucleotide sequence of claim 34.

30 37. An isolated nucleic acid comprising a nucleotide sequence encoding the Serrate protein of claim 2.

38. An isolated nucleic acid comprising the Serrate coding sequence contained in plasmid pBS39 as 35 deposited with the ATCC and assigned accession number 97068.

39. An isolated human nucleic acid hybridizable to plasmid SerFL or the Serrate sequence therein as deposited with the ATCC and assigned accession number 68876.
- 5           40. An isolated nucleic acid comprising the Serrate coding sequence contained in plasmid pBS3-2 as deposited with the ATCC and assigned accession number \_\_\_\_\_.
- 10           41. An isolated nucleic acid comprising the Serrate coding sequence contained in plasmid pBS15 as deposited with the ATCC and assigned accession number \_\_\_\_\_.
- 15           42. An isolated nucleic acid comprising a nucleotide sequence encoding a protein, said protein comprising amino acid numbers 1 - 1257 of SEQ ID NO:4.
- 20           43. An isolated nucleic acid comprising a fragment of a vertebrate Serrate gene consisting of at least 8 nucleotides.
- 25           44. An isolated nucleic acid comprising a nucleotide sequence encoding the fragment of claim 14, 15, 16 or 25.
- 30           45. The nucleic acid of claim 44 in which the fragment is a fragment of a human Serrate protein.
- 35           46. An isolated nucleic acid comprising a nucleotide sequence encoding the fragment of claim 12.
47. An isolated nucleic acid comprising a nucleotide sequence encoding a protein, said protein comprising amino acid numbers 30 - 1218 of SEQ ID NO:2.
48. An isolated nucleic acid comprising a nucleotide sequence encoding the protein of claim 21.

49. A recombinant cell containing the nucleic acid of claim 34, 37 or 43.

50. A recombinant cell containing the nucleic acid of claim 38, 40 or 41.

51. A method of producing a Serrate protein comprising growing a recombinant cell containing the nucleic acid of claim 34 or 37 such that the encoded Serrate protein is expressed by the cell, and recovering the expressed Serrate protein.

52. A method of producing a Serrate protein comprising growing a recombinant cell containing the nucleic acid of claim 38, 40 or 41 such that the encoded Serrate protein is expressed by the cell, and recovering the expressed Serrate protein.

53. A method of producing a Serrate protein comprising growing a recombinant cell containing the nucleic acid of claim 45 such that the encoded protein is expressed by the cell, and recovering the expressed protein.

54. A method of producing a protein comprising a fragment of a Serrate protein, which method comprises growing a recombinant cell containing the nucleic acid of claim 46 such that the encoded protein is expressed by the cell, and recovering the expressed protein.

55. The product of the process of claim 51.

56. The product of the process of claim 52.

57. The product of the process of claim 53.

58. The product of the process of claim 54.

59. A pharmaceutical composition comprising a therapeutically effective amount of a vertebrate Serrate protein; and a pharmaceutically acceptable carrier.

5 60. The composition of claim 59 in which the Serrate protein is a human Serrate protein.

61. A pharmaceutical composition comprising a therapeutically effective amount of the fragment of claim 14,  
10 15, 16 or 25; and a pharmaceutically acceptable carrier.

62. A pharmaceutical composition comprising a therapeutically effective amount of the fragment of claim 12; and a pharmaceutically acceptable carrier.

15

63. A pharmaceutical composition comprising a therapeutically effective amount of a molecule comprising a fragment of a vertebrate Serrate protein, which derivative or analog is characterized by the ability to bind to a Notch  
20 protein or to a molecule comprising the epidermal growth factor-like repeats 11 and 12 of a Notch protein.

64. A pharmaceutical composition comprising a therapeutically effective amount of the nucleic acid of claim  
25 34, 36 or 37; and a pharmaceutically acceptable carrier.

65. A pharmaceutical composition comprising a therapeutically effective amount of the nucleic acid of claim 44; and a pharmaceutically acceptable carrier.

30

66. A pharmaceutical composition comprising a therapeutically effective amount of the nucleic acid of claim 46; and a pharmaceutically acceptable carrier.

35 67. A pharmaceutical composition comprising a therapeutically effective amount of the antibody of claim 30; and a pharmaceutically acceptable carrier.

68. A pharmaceutical composition comprising a therapeutically effective amount of a fragment or derivative of the antibody of claim 30 containing the binding domain of the antibody; and a pharmaceutically acceptable carrier.

5

69. A method of treating or preventing a disease or disorder in a subject comprising administering to a subject in which such treatment or prevention is desired a therapeutically effective amount of a vertebrate Serrate  
10 protein or derivative thereof which is able to bind to a Notch protein.

70. The method according to claim 69 in which the disease or disorder is a malignancy characterized by  
15 increased Notch activity or increased expression of a Notch protein or of a Notch derivative capable of being bound by an anti-Notch antibody, relative to said Notch activity or expression in an analogous non-malignant sample.

20 71. The method according to claim 69 in which the disease or disorder is selected from the group consisting of cervical cancer, breast cancer, colon cancer, melanoma, seminoma, and lung cancer.

25 72. The method according to claim 69 in which the subject is a human.

73. The method according to claim 69 in which the Serrate protein is a human Serrate protein.

30

74. A method of treating or preventing a disease or disorder in a subject comprising administering to a subject in which such treatment or prevention is desired a therapeutically effective amount of a molecule, in which the  
35 molecule is an oligonucleotide which (a) comprises ten nucleotides; (b) comprises a sequence absolutely complementary to an at least ten nucleotide portion of an RNA

transcript specific to a vertebrate Serrate gene; and (c) is hybridizable to the RNA transcript.

75. A method of treating or preventing a disease  
5 or disorder in a subject comprising administering to a  
subject in which such treatment or prevention is desired an  
effective amount of the nucleic acid of claim 34, 37 or 46.

76. A method of treating or preventing a disease  
10 or disorder in a subject comprising administering to a  
subject in which such treatment or prevention is desired an  
effective amount of the antibody of claim 32.

77. The method according to claim 73 in which the  
15 disease or disorder is a disease or disorder of the central  
nervous system.

78. An isolated oligonucleotide comprising ten  
nucleotides, and comprising a sequence absolutely  
20 complementary to an at least ten nucleotide portion of an RNA  
transcript specific to a vertebrate Serrate gene, which  
oligonucleotide is hybridizable to the RNA transcript.

79. A pharmaceutical composition comprising the  
25 oligonucleotide of claim 78; and a pharmaceutically  
acceptable carrier.

80. A method of inhibiting the expression of a  
nucleic acid sequence encoding a Serrate protein in a cell  
30 comprising providing the cell with an effective amount of the  
oligonucleotide of claim 78.

81. A method of diagnosing a disease or disorder  
characterized by an aberrant level of Notch-Serrate protein  
35 binding activity in a patient, comprising measuring the  
ability of a Notch protein in a sample derived from the  
patient to bind to a vertebrate Serrate protein, in which an

increase or decrease in the ability of the Notch protein to bind to the Serrate protein, relative to the ability found in an analogous sample from a normal individual, indicates the presence of the disease or disorder in the patient.

5

82. A method of diagnosing a disease or disorder characterized by an aberrant level of Serrate protein in a patient, comprising measuring the levels of a vertebrate Serrate protein in a sample derived from the patient, in  
10 which an increase or decrease in the levels of the Serrate protein, relative to the levels of the Serrate protein found in an analogous sample from a normal individual, indicates the presence of the disease or disorder in the patient.

15

20

25

30

35

1/20

10	20	30	40	50	60
GAATTCCT	CCCCCTTT	TCCATGCAGC	TGATCTAAAA	GGGAATAAAA	GGCTGCGCAT
70	80	90	100	110	120
AATCATAATA	ATAAAAGAAG	GGGAGCGCGA	GAGAAGGAAA	GAAAGCCGGG	AGGTGGAAGA
130	140	150	160	170	180
GGAGGGGGAG	CGTCTCAAAG	AAGCGATCAG	AATAATAAAA	GGAGGCCGGG	CTCTTTGCCT
190	200	210	220	230	240
TCTGGAAGGG	GCCGCTCTTG	AAAGGGCTTT	TGAAAAGTGG	TGTTGTTTTTC	CAGTCGTGCA
250	260	270	280	290	300
TGCTCCAATC	GGCGGAGTAT	ATTAGAGCCG	GGACGCGGCC	GCAGGGGCAG	CGGCGACGGC
310	320	330	340	350	360
AGCACCGGCG	GCAGCACGAG	CGCGAACAGC	AGCGGCGGCG	TCCCGAGTGC	CCGCGGCGGC
370	380	390	400	410	420
GCGCGCAGCG	ATGCGTTCCC	CACGGACACG	CGGCCGTGCC	GGGCGCCCCC	TAAGCCTCCT
	M R S	P R T R	G R S	G R P	L S L L>
430	440	450	460	470	480
GCTCGCCCTG	CTCTGTGCCC	TGCGAGCCAA	GGTGTGTGGG	GCCTCGGGTC	AGTTCGAGTT
L A L	L C A	L R A K	V C G	A S G	Q F E L>
490	500	510	520	530	540
GGAGATCCTG	TCCATGCAGA	ACGTGAACGG	GGAGCTGCAG	AACGGGAACT	GCTGCGGCGG
E I L	S M Q	N V N G	E L Q	N G N	C C G G>
550	560	570	580	590	600
CGCCCGGAAC	CCGGGAGACC	GCAAGTGCAC	CCGCGACGAG	TGTGACACAT	ACTTCAAAGT
A R N	P G D	R K C T	R D E	C D T	Y F K V>
610	620	630	640	650	660
GTGCCTCAAG	GAGTATCAGT	CCCGCGTCAC	GGCCGGGGGG	CCCTGCAGCT	TCGGCTCAGG
C L K	E Y Q	S R V T	A G G	P C S	F G S G>
670	680	690	700	710	720
GTCCACGCCT	GTCATCGGGG	GCAACACCTT	CAACCTCAAG	GCCAGCCGCG	GCAACGACCC
S T P	V I G	G N T F	N L K	A S P	G N D P>
730	740	750	760	770	780
GAACCGCATC	GTGCTGCCTT	TCAGTTTCGC	CTGGCCGAGG	TCCTATACGT	TGCTTGTGGA
N R I	V L P	F S F A	W P R	S Y T	L L V E>
790	800	810	820	830	840
GGCGTGGGAT	TCCAGTAATG	ACACCGTTCA	ACCTGACAGT	ATTATTGAAA	AGGCTTCTCA
A W D	S S N	D T V Q	P D S	I I E	K A S H>
850	860	870	880	890	900
CTCGGGCATG	ATCAACCCCA	GCCGGCAGTG	GCAGACGCTG	AAGCAGAACA	CGGGCGTTGC
S G M	I N P	S R Q W	Q T L	K Q N	T G V A>

FIG. 1A



2/20

910	920	930	940	950	960
CCACTTTGAG	TATCAGATCC	GCGTGACCTG	TGATGACTAC	TACTATGGCT	TTGGCTGTAA
H F E Y Q I	R V T C	D D Y	Y Y G	F G C N>	
970	980	990	1000	1010	1020
TAAGTTCTGC	CGCCCCAGAG	ATGACTTCTT	TGGACACTAT	GCCTGTGACC	AGAATGGCAA
K F C R P R	D D F F	G H Y	A C D	Q N G N>	
1030	1040	1050	1060	1070	1080
CAAAACTTGC	ATGGAAGGCT	GGATGGGCCC	CGAATGTAAC	AGAGCTATTT	GCCGACAAGG
K T C M E G	W M G P	E C N	R A I	C R Q G>	
1090	1100	1110	1120	1130	1140
CTGCAGTCCT	AAGCATGGGT	CTTGCAAAC	CCCAGGTGAC	TGCAGGTGCC	AGTACGGCTG
C S P K H G	S C K L	P G D	C R C	Q Y G W>	
1150	1160	1170	1180	1190	1200
GCAAGGCCTG	TACTGTGATA	AGTGCATCCC	ACACCCGGGA	TGCGTCCACG	GCATCTGTAA
Q G L Y C D	K C I P	H P G	C V H	G I C N>	
1210	1220	1230	1240	1250	1260
TGAGCCCTGG	CAGTGCCTCT	GTGAGACCAA	CTGGGGCGGC	CAGCTCTGTG	ACAAAGATCT
E P W Q C L	C E T N	W G G	Q L C	D K D L>	
1270	1280	1290	1300	1310	1320
CAATTACTGT	GGGACTCATC	AGCCGTGTCT	CAACGGGGGA	ACTTGTAGCA	ACACAGGCCC
N Y C G T H	Q P C L	N G G	T C S	N T G P>	
1330	1340	1350	1360	1370	1380
TGACAAATAT	CAGTGTTCCT	GCCCTGAGGG	GTATTGAGGA	CCCAACTGTG	AAATTGCTGA
D K Y Q C S	C P E G	Y S G	P N C	E I A E>	
1390	1400	1410	1420	1430	1440
GCACGCCTGC	CTCTCTGATC	CCTGTCACAA	CAGAGGCAGC	TGTAAGGAGA	CCTCCCTGGG
H A C L S D	P C H N	R G S	C K E	T S L G>	
1450	1460	1470	1480	1490	1500
CTTTGAGTGT	GAGTGTTCCT	CAGGCTGGAC	CGGCCCCACA	TGCTCTACAA	ACATTGATGA
F E C E C S	P G W T	G P T	C S T	N I D D>	
1510	1520	1530	1540	1550	1560
CTGTTCTCCT	AATAACTGTT	CCCACGGGGG	CACCTGCCAG	GACCTGGTTA	ACGGATTTAA
C S P N N C	S H G G	T C Q	D L V	N G F K>	
1570	1580	1590	1600	1610	1620
GTGTGTGTGC	CCCCACAGT	GGACTGGGAA	AACGTGCCAG	TTAGATGCAA	ATGAATGTGA
C V C P P Q	W T G K	T C Q	L D A	N E C E>	
1630	1640	1650	1660	1670	1680
GGCCAAACCT	TGTGTAAACG	CCAAATCCTG	TAAGAATCTC	ATTGCCAGCT	ACTACTGCGA
A K P C V N	A K S C	K N L	I A S	Y Y C D>	

FIG. 1B

SUBSTITUTE SHEET (RULE 26)

3/20

1690	1700	1710	1720	1730	1740
CTGTCTTCCC	GGCTGGATGG	GTCAGAATTG	TGACATAAAT	ATTAATGACT	GCCTTGGCCA
C L P	G W M	G Q N C	D I N	I N D	C L G Q>
1750	1760	1770	1780	1790	1800
GTGTCAGAAT	GACGCCTCCT	GTCGGGATTT	GGTTAATGGT	TATCGCTGTA	TCTGTCCACC
C Q N	D A S	C R D L	V N G	Y R C	I C P P>
1810	1820	1830	1840	1850	1860
TGGCTATGCA	GGCGATCACT	GTGAGAGAGA	CATCGATGAA	TGTGCCAGCA	ACCCCTGTTT
G Y A	G D H	C E R D	I D E	C A S	N P C L>
1870	1880	1890	1900	1910	1920
GAATGGGGGT	CACTGTCAGA	ATGAAATCAA	CAGATTCCAG	TGTCTGTGTC	CCACTGGTTT
N G G	H C Q	N E I N	R F Q	C L C	P T G F>
1930	1940	1950	1960	1970	1980
CTCTGGAAAC	CTCTGTCAGC	TGGACATCGA	TTATTGTGAG	CCTAATCCCT	GCCAGAACGG
S G N	L C Q	L D I D	Y C E	P N P	C Q N G>
1990	2000	2010	2020	2030	2040
TGCCCAGTGC	TACAACCGTG	CCAGTGACTA	TTTCTGCAAG	TGCCCCGAGG	ACTATGAGGG
A Q C	Y N R	A S D Y	F C K	C P E	D Y E G>
2050	2060	2070	2080	2090	2100
CAAGAACTGC	TCACACCTGA	AAGACCACTG	CCGCACGACC	CCCTGTGAAG	TGATTGACAG
K N C	S H L	K D H C	R T T	P C E	V I D S>
2110	2120	2130	2140	2150	2160
CTGCACAGTG	GCCATGGCTT	CCAACGACAC	ACCTGAAGGG	GTGCGGTATA	TTTCCTCCAA
C T V	A M A	S N D T	P E G	V R Y	I S S N>
2170	2180	2190	2200	2210	2220
CGTCTGTGGT	CCTCACGGGA	AGTGCAAGAG	TCAGTCGGGA	GGCAAATTCA	CCTGTGACTG
V C G	P H G	K C K S	Q S G	G K F	T C D C>
2230	2240	2250	2260	2270	2280
TAACAAAGGC	TTCACGGGAA	CATACTGCCA	TGAAAAATATT	AATGACTGTG	AGAGCAACCC
N K G	F T G	T Y C H	E N I	N D C	E S N P>
2290	2300	2310	2320	2330	2340
TTGTAGAAAC	GGTGGCACTT	GCATCGATGG	TGTCAACTCC	TACAAGTGCA	TCTGTAGTGA
C R N	G G T	C I D G	V N S	Y K C	I C S D>
2350	2360	2370	2380	2390	2400
CGGCTGGGAG	GGGGCCTACT	GTGAAACCAA	TATTAATGAC	TGCAGCCAGA	ACCCCTGCCA
G W E	G A Y	C E T N	I N D	C S Q	N P C H>
2410	2420	2430	2440	2450	2460
CAATGGGGGC	ACGTGTCGCG	ACCTGGTCAA	TGACTTCTAC	TGTGACTGTA	AAAATGGGTG
N G G	T C R	D L V N	D F Y	C D C	K N G W>

FIG. 1C

4/20

2470 2480 2490 2500 2510 2520  
GAAAGGAAAG ACCTGCCACT CACGTGACAG TCAGTGTGAT GAGGCCACGT GCAACAACGG  
K G K T C H S R D S Q C D E A T C N N G>  
2530 2540 2550 2560 2570 2580  
TGGCACCTGC TATGATGAGG GGGATGCTTT TAAGTGCATG TGTCTGGCG GCTGGGAAGG  
G T C Y D E G D A F K C M C P G G W E G>  
2590 2600 2610 2620 2630 2640  
AACAACTGT AACATAGCCC GAAACAGTAG CTGCCTGCCC AACCCCTGCC ATAATGGGGG  
T T C N I A R N S S C L P N P C H N G G>  
2650 2660 2670 2680 2690 2700  
CACATGTGTG GTCAACGGCG AGTCCTTTAC GTGCGTCTGC AAGGAAGGCT GGGAGGGGCC  
T C V V N G E S F T C V C K E G W E G P>  
2710 2720 2730 2740 2750 2760  
CATCTGTGCT CAGAATACCA ATGACTGCAG CCCTCATCCC TGTTACAACA GCGGCACCTG  
I C A Q N T N D C S P H P C Y N S G T C>  
2770 2780 2790 2800 2810 2820  
TGTGGATGGA GACAACTGGT ACCGGTGCGA ATGTGCCCCG GGTTTGTCTG GGCCCCACTG  
V D G D N W Y R C E C A P G F A G P D C>  
2830 2840 2850 2860 2870 2880  
CAGAATAAAC ATCAATGAAT GCCAGTCTTC ACCTTGTGCC TTTGGAGCGA CCTGTGTGGA  
R I N I N E C Q S S P C A F G A T C V D>  
2890 2900 2910 2920 2930 2940  
TGAGATCAAT GGCTACCGGT GTGTCTGCCC TCCAGGGCAC AGTGGTGCCA AGTGCCAGGA  
E I N G Y R C V C P P G H S G A K C Q E>  
2950 2960 2970 2980 2990 3000  
AGTTTCAGGG AGACCTTGCA TCACCATGGG GAGTGTGATA CCAGATGGGG CCAAATGGGA  
V S G R P C I T M G S V I P D G A K W D>  
3010 3020 3030 3040 3050 3060  
TGATGACTGT AATACCTGCC AGTGCCTGAA TGGACGGATC GCCTGCTCAA AGGTCTGGTG  
D D C N T C Q C L N G R I A C S K V W C>  
3070 3080 3090 3100 3110 3120  
TGGCCCTCGA CCTTGCCTGC TCCACAAAGG GCACAGCGAG TGCCCCAGCG GGCAGAGCTG  
G P R P C L L H K G H S E C P S G Q S C>  
3130 3140 3150 3160 3170 3180  
CATCCCCATC CTGGACGACC AGTGCTTCGT CCACCCCTGC ACTGGTGTGG GCGAGTGTG  
I P I L D D Q C F V H P C T G V G E C R>  
3190 3200 3210 3220 3230 3240  
GTCTTCCAGT CTCCAGCCGG TGAAGACAAA GTGCACCTCT GACTCCTATT ACCAGGATAA  
S S S L Q P V K T K C T S D S Y Y Q D N>  
3250 3260 3270 3280 3290 3300  
CTGTGCGAAC ATCACATTTA CCTTTAACAA GGAGATGATG TCACCAGGTC TTACTACGGA  
C A N I T F T F N K E M M S P G L T T E>

FIG. 1D

SUBSTITUTE SHEET (RULE 26)

5/20

3310 3320 3330 3340 3350 3360  
GCACATTTGC AGTGAATTGA GGAATTTGAA TATTTTGAAG AATGTTTCCG CTGAATATTC  
H I C S E L R N L N I L K N V S A E Y S>  
3370 3380 3390 3400 3410 3420  
AATCTACATC GCTTGCAGAGC CTTCCCCTTC AGCGAACAAT GAAATACATG TGGCCATTTC  
I Y I A C E P S P S A N N E I H V A I S>  
3430 3440 3450 3460 3470 3480  
TGCTGAAGAT ATACGGGATG ATGGGAACCC GATCAAGGAA ATCACTGACA AAATAATCGA  
A E D I R D D G N P I K E I T D K I I D>  
3490 3500 3510 3520 3530 3540  
TCTTGTTACT AAACGTGATG GAAACAGCTC GCTGATTGCT GCCGTTGAAG AAGTAAGAGT  
L V T K R D G N S S L I A A V E E V R V>  
3550 3560 3570 3580 3590 3600  
TCAGAGGCGG CCTCTGAAGA ACAGAACAGA TTTCCTTGTT CCCTTGCTGA GCTCTGTCTT  
Q R R P L K N R T D F L V P L L S S V L>  
3610 3620 3630 3640 3650 3660  
AACTGTGGCT TGGATCTGTT GCTTGGTGAC GGCCTTCTAC TGGTGCCTGC GGAAGCGGCG  
T V A W I C C L V T A F Y W C L R K R R>  
3670 3680 3690 3700 3710 3720  
GAAGCCGGGC AGCCACACAC ACTCAGCCTC TGAGGACAAC ACCACCAACA ACGTGCGGGA  
K P G S H T H S A S E D N T T N N V R E>  
3730 3740 3750 3760 3770 3780  
GCAGCTGAAC CAGATCAAAA ACCCCATTGA GAAACATGGG GCCAACACGG TCCCCATCAA  
Q L N Q I K N P I E K H G A N T V P I K>  
3790 3800 3810 3820 3830 3840  
GGATTACGAG AACAAGAAGT CCAAAATGTC TAAAATAAGG ACACACAATT CTGAAGTAGA  
D Y E N K N S K M S K I R T H N S E V E>  
3850 3860 3870 3880 3890 3900  
AGAGGACGAC ATGGACAAAC ACCAGCAGAA AGCCCGGTTT GCCAAGCAGC CGGCGTACAC  
E D D M D K H Q Q K A R F A K Q P A Y T>  
3910 3920 3930 3940 3950 3960  
GCTGGTAGAC AGAGAAGAGA AGCCCCCAA CGGCACGCCG AAAAAACACC CAAACTGGAC  
L V D R E E K P P N G T P T K H P N W T>  
3970 3980 3990 4000 4010 4020  
AAACAAACAG GACAACAGAG ACTTGGAAG TGCCCAGAGC TTAAACCGAA TGGAGTACAT  
N K Q D N R D L E S A Q S L N F M E Y I>  
4030 4040 4050 4060 4070 4080  
CGTATAGCAG ACCGCGGGCA CTGCCGCCGC TAGGTAGAGT CTGAGGGCTT GTAGTTCCTT  
V >

FIG. 1E

6/20

4090	4100	4110	4120	4130	4140
AAACTGTCGT	GTCATACTCG	AGTCTGAGGC	CGTTGCTGAC	TTAGAATCCC	TGTGTTAATT
4150	4160	4170	4180	4190	4200
TAGTTTGACA	AGCTGGCTTA	CACTGGCAAT	GGTAGTTCTG	TGGTTGGCTG	GGAAATCGAG
4210	4220	4230	4240	4250	4260
TGGCGCATCT	CACAGCTATG	CAAAAAGCTA	GTCAACAGTA	CCCCTGGTTG	TGTGTCCCCT
4270	4280	4290	4300	4310	4320
TGCAGCCGAC	ACGGTCTCGG	ATCAGGCTCC	CAGGAGCTGC	CCAGCCCCCT	GGTACTTTGA
4330	4340	4350	4360	4370	4380
GCTCCCACTT	CTGCCAGATG	TCTAATGGTG	ATGCAGTCTT	AGATCATAGT	TTTATTTATA
4390	4400	4410	4420	4430	4440
TTTATTGACT	CTTGAGTTGT	TTTTGTATAT	TGGTTTTATG	ATGACGTACA	AGTAGTTCTG
4450	4460	4470	4480	4490	4500
TATTTGAAAG	TGCCTTTGCA	GCTCAGAAC	ACAGCAACGA	TCACAAATGA	CTTTATTATT
4510	4520	4530	4540	4550	4560
TATTTTTTTT	AATTGTATTT	TTGTTGTTGG	GGGAGGGGAG	ACTTTGATGT	CAGCAGTTGC
4570	4580	4590	4600	4610	4620
TGGTAAATG	AAGAATTTAA	AGAAAAATG	TCCAAAAGTA	GAACTTTGTA	TAGTTATGTA
4630	4640	4650	4660	4670	4680
AATAATTCTT	TTTTATTAAT	CACTGTGTAT	ATTTGATTTA	TTAACTTAAT	AATCAAGAGC
4690	4700	4710	4720	4730	4740
CTTAAACAT	CATTCTTTT	TATTTATATG	TATGTGTTTA	GAATTGAAGG	TTTTTGATAG
4750	4760	4770	4780	4790	4800
CATTGTAAGC	GTATGGCTTT	ATTTTTTTGA	ACTCTTCTCA	TTACTTGTTG	CCTATAAGCC
4810	4820	4830	4840	4850	4860
AAAAAGGAAA	GGGTGTTTTG	AAAATAGTTT	ATTTTAAAC	AATAGGATGG	GCTACACGTA
4870	4880	4890	4900	4910	4920
CATAGGTAAG	TAATAGCACC	GTACTGGTTA	TGATGATGAA	AATAACTGGA	AACTTGAAAG
4930	4940	4950	4960	4970	4980
CTTGTGGTAA	TGGCAGATAA	AGATGGTTCA	CCTGGGAAAT	TAAACTTGA	ATGGTTGTAC
4990	5000	5010	5020	5030	5040
AGAAAAGCAC	AGAGTGGAAT	GCACATCAAT	GACAGTAAGG	GAGTTAGTTC	TAGGAACAGC
5050	5060	5070	5080	5090	5100
TCCTGAACAG	TAAGATTCCC	GCAATAGTCT	CCGCTCGTT	CGTCTATGGT	ATGCATCCCA
5110	5120	5130	5140	5150	5160
TTCATTTTCT	TCTTCTGATT	ATTGTCATCT	TTCCCTTTGC	CAAATGGGCA	GTTATTGTTT
5170	5180	5190	5200	5210	5220
CAGGGAGAGA	AGCTGCTCAT	TGGCCAATCA	TTCTGGTGTG	CAGTGCTCCA	TCGGATTCTA
5230	5240	5250	5260	5270	5280
CATGTCCAAC	AAGGCATGTC	TGGATGATGC	AATGTCTGTC	TGACCCCCGG	AATTCCGTGC

FIG. 1F

SUBSTITUTE SHEET (RULE 26)

7/20

5290	5300	5310	5320	5330	5340
AGAGACAACA	TTCTAGACAG	ATATACACTT	TTTATTATTA	ACAAACTTTG	GCCACAACCT
5350	5360	5370	5380	5390	5400
TTGATGTATA	AATTGCCGGA	TTTCCCCAGT	CCTTTCATTG	TGGCTTTGGA	CAGGAGCAGG
5410	5420	5430	5440	5450	5460
CTCACTTGTC	TGCTTCAGGC	TGCCTTTCTC	TTGGGTTGCA	CCTCAGTTCT	TACTTATTTA
5470	5480	5490	5500	5510	5520
TTTATTTTGA	GTGGAGCATA	GGGGCCTCTT	CCAAAATGGG	TAGAGCTCAG	GGGCTTTCTT
5530	5540	5550	5560	5570	5580
ATTGAAATGG	TCACATGATA	AAAACGGGCT	GAAAAAGGAG	AGTTCCAGGA	GAAAAGCCCCA
5590	5600	5610	5620	5630	5640
GAAAAGGCCC	CTCCTCAGAA	GACAGCCTTT	AAGCCTCTTG	CTTACTGAAG	GAAGCCCCAC
5650	5660	5670	5680	5690	5700
CTTCTAGCAC	TGAGGCCGGG	TCTGATCTTC	CAGAGGAGTT	GGAGGAGTCC	ATGAGAATGG
5710	5720	5730	5740	5750	5760
CCACCATTCT	TGCTTGCTGC	TGCTGATGTT	GCAGTTTTGA	GAGAACAGCG	GGATCCTTGT
5770	5780	5790	5800	5810	5820
TGTCCTCTAG	AGACTTGAGT	CTGTCACTGA	CATTTTTTCA	GTTCCCTTGC	TCATAGACCA
5830	5840	5850	5860	5870	5880
TACGAGGAAT	TAGTGATGTG	TCAGTTGAGA	GTTCAACAATC	TCATTGTTCA	TTTAATTCAC
5890	5900	5910	5920	5930	5940
TTTAAAGTTG	TCAATTTCTG	TGTGAGTAAC	CTGTAAAAGA	CACCTTTCCA	GAAGAGTTTT
5950	5960	5970	5980	5990	6000
GCCGTCTGTT	TGAAAAAATA	ATCTTTATAA	ACTTTCCTAA	GTATCTGGAT	TTGGATTCTT
6010	6020	6030	6040	6050	6060
TATTTGGAGA	GAAAATGTAC	CCTGTCTCCA	CCAAAATAC	AAAAATTAGC	CAGGCTTGGT
6070	6080	6090	6100	6110	6120
GGTGACACAC	GGTAATCCCA	GCAACTCTGG	AGACTAAGGC	AGGAAGAATC	GCTTGACCCA
6130	6140	6150	6160	6170	6180
GGAGGGTCGA	GGCTACAATG	AGTTGAAACC	GCGCCACTGC	ACTCCAGCCT	GGGCGACAGT
6190	6200	6210	6220	6230	6240
GCGAGGCCCT	GTCTCAAAAA	TAAAATAAAA	TAAATAAATA	AATTAGCCAG	ATACTGTGTG
6250	6260	6270	6280	6290	6300
CACGCCTGCA	GTCCCAGCTA	TTCTGGAAGC	TGAGGTGGGA	AGATGGTTAA	GCCTGAGAGG
6310	6320	6330	6340	6350	6360
ACAAAGCTGC	AGTGAGTCAT	GTTTGCATCA	CTGCACTCCA	GCCTGGGTGA	CAGAGCAAGA
6370	6380	6390	6400	6410	6420
CCCTGTCTAA	AAAACAAAAA	CAGGCCGGGT	GTGGTGGCTC	ATGCCTGCCA	TCCCAGTGCT
6430	6440	6450	6460		
TTGGGAGGCA	GAGGTTGGCA	TAATCCCAGC	GCTCTGGGAA	TTCC	

FIG. 1G

SUBSTITUTE SHEET (RULE 26)

GGCCGGGGCC	GGGCGGGCGG	GTGCGGGGGG	CAATGCGGGC	GCAGGGCCGG	GGGCGCCTTC	60										
CCCGGCGGCT	GCTGCTGCTG	CTGGCGCTCT	GGGTGCAGGC	GGCGCGGCC	ATGGGCTATT	120										
TCGAGCTGCA	GCTGAGCGCG	CTGCGGAACG	TGAACGGGGA	GCTGCTGAGC	GGCGCCTGCT	180										
GTGACGGCGA	CGGCCGGACA	ACGCGCGCGG	GGGGCTGCGG	CCACGACGAG	TGCGACACCG	240										
CTCCTTTACC	CTCATCGTGG	AGGCCTGGGA	CTGGGACAAC	GATACCACCC	CGAATGAGGA	300										
GCTGCTGATC	GAGCGAGTGT	CGCATGCCGG	C	ATG	ATC	AAC	CCG	GAG	GAC	CGC	352					
				Met	Ile	Asn	Pro	Glu	Asp	Arg						
				1				5								
TGG	AAG	AGC	CTG	CAC	TTC	AGC	GGC	CAC	GTG	GCG	CAC	CTG	GAG	CTG	CAG	400
Trp	Lys	Ser	Leu	His	Phe	Ser	Gly	His	Val	Ala	His	Leu	Glu	Leu	Gln	
		10					15					20				
ATC	CGC	GTG	CGC	TGC	GAC	GAG	AAC	TAC	TAC	AGC	GCC	ACT	TGC	AAC	AAG	448
Ile	Arg	Val	Arg	Cys	Asp	Glu	Asn	Tyr	Tyr	Ser	Ala	Thr	Cys	Asn	Lys	
	25					30					35					
TTC	TGC	CGG	CCC	CGC	AAT	GAC	TTT	TTC	GGC	CAC	TAC	ACC	TGC	GAC	CAG	496
Phe	Cys	Arg	Pro	Arg	Asn	Asp	Phe	Phe	Gly	His	Tyr	Thr	Cys	Asp	Gln	
	40				45				50						55	
TAC	GGC	AAC	AAG	GCC	TGC	ATG	GAC	GGC	TGG	ATG	GGC	AAG	GAG	TGC	AAG	544
Tyr	Gly	Asn	Lys	Ala	Cys	Met	Asp	Gly	Trp	Met	Gly	Lys	Glu	Cys	Lys	
			60					65					70			
GAA	GCT	GTG	TGT	AAA	CAA	GGG	TGT	AAT	TTG	CTC	CAC	GGG	GGA	TGC	ACC	592
Glu	Ala	Val	Cys	Lys	Gln	Gly	Cys	Asn	Leu	Leu	His	Gly	Gly	Cys	Thr	
			75					80					85			
GTG	CCT	GGG	GAG	TGC	AGG	TGC	AGC	TAC	GGC	TGG	CAA	GGG	AGG	TTC	TGC	640
Val	Pro	Gly	Glu	Cys	Arg	Cys	Ser	Tyr	Gly	Trp	Gln	Gly	Arg	Phe	Cys	
		90					95					100				
GAT	GAG	TGT	GTC	CCC	TAC	CCC	GGC	TGC	GTG	CAT	GGC	AGT	TGT	GTG	GAG	688
Asp	Glu	Cys	Val	Pro	Tyr	Pro	Gly	Cys	Val	His	Gly	Ser	Cys	Val	Glu	
	105					110					115					
CCC	TGG	CAG	TGC	AAC	TGT	GAG	ACC	AAC	TGG	GGC	GGC	CTG	CTC	TGT	GAC	736
Pro	Trp	Gln	Cys	Asn	Cys	Glu	Thr	Asn	Trp	Gly	Gly	Leu	Leu	Cys	Asp	
	120				125				130						135	
AAA	GAC	CTG	AAC	TAC	TGT	GGC	AGC	CAC	CAC	CCC	TGC	ACC	AAC	GGA	GGC	784
Lys	Asp	Leu	Asn	Tyr	Cys	Gly	Ser	His	His	Pro	Cys	Thr	Asn	Gly	Gly	
			140					145						150		

FIG. 2A

**SUBSTITUTE SHEET (RULE 26)**

9/20

ACG TGC ATC AAC GCC GAG CCT GAC CAG TAC CGC TGC ACC TGC CCT GAC 832  
 Thr Cys Ile Asn Ala Glu Pro Asp Gln Tyr Arg Cys Thr Cys Pro Asp  
 155 160 165  
 GGC TAC TCG GGC AGG AAC TGT GAG AAG GCT GAG CAC GCC TGC ACC TCC 880  
 Gly Tyr Ser Gly Arg Asn Cys Glu Lys Ala Glu His Ala Cys Thr Ser  
 170 175 180  
 AAC CCG TGT GCC AAC GGG GGC TCT TGC CAT GAG GTG CCG TCC GGC TTC 928  
 Asn Pro Cys Ala Asn Gly Gly Ser Cys His Glu Val Pro Ser Gly Phe  
 185 190 195  
 GAA TGC CAC TGC CCA TCG GGC TGG AGC GGG CCC ACC TGT GCC CTT GAC 976  
 Glu Cys His Cys Pro Ser Gly Trp Ser Gly Pro Thr Cys Ala Leu Asp  
 200 205 210 215  
 ATC GAT GAG TGT GCT TCG AAC CCG TGT GCG GCC GGT GGC ACC TGT GTG 1024  
 Ile Asp Glu Cys Ala Ser Asn Pro Cys Ala Ala Gly Gly Thr Cys Val  
 220 225 230  
 GAC CAG GTG GAC GGC TTT GAG TGC ATC TGC CCC GAG CAG TGG GTG GGC 1072  
 Asp Gln Val Asp Gly Phe Glu Cys Ile Cys Pro Glu Gln Trp Val Gly  
 235 240 245  
 GCC ACC TGC CAG CTG GAC GCC AAT GAG TGT GAA GGG AAG CCA TGC CTT 1120  
 Ala Thr Cys Gln Leu Asp Ala Asn Glu Cys Glu Gly Lys Pro Cys Leu  
 250 255 260  
 AAC GCT TTT TCT TGC AAA AAC CTG ATT GGC GGC TAT TAC TGT GAT TGC 1168  
 Asn Ala Phe Ser Cys Lys Asn Leu Ile Gly Gly Tyr Tyr Cys Asp Cys  
 265 270 275  
 ATC CCG GGC TGG AAG GGC ATC AAC TGC CAT ATC AAC GTC AAC GAC TGT 1216  
 Ile Pro Gly Trp Lys Gly Ile Asn Cys His Ile Asn Val Asn Asp Cys  
 280 285 290 295  
 CGC GGG CAG TGT CAG CAT GGG GGC ACC TGC AAG GAC CTG GTG AAC GGG 1264  
 Arg Gly Gln Cys Gln His Gly Gly Thr Cys Lys Asp Leu Val Asn Gly  
 300 305 310  
 TAC CAG TGT GTG TGC CCA CGG GGC TTC GGA GGC CGG CAT TGC GAG CTG 1312  
 Tyr Gln Cys Val Cys Pro Arg Gly Phe Gly Gly Arg His Cys Glu Leu  
 315 320 325  
 GAA CGA GAC AAG TGT GCC AGC AGC CCC TGC CAC AGC GGC GGC CTC TGC 1360  
 Glu Arg Asp Lys Cys Ala Ser Ser Pro Cys His Ser Gly Gly Leu Cys  
 330 335 340  
 GAG GAC CTG GCC GAC GGC TTC CAC TGC CAC TGC CCC CAG GGC TTC TCC 1408  
 Glu Asp Leu Ala Asp Gly Phe His Cys His Cys Pro Gln Gly Phe Ser  
 345 350 355

FIG. 2B

SUBSTITUTE SHEET (RULE 26)



10/20

GGG	CCT	CTC	TGT	GAG	GTG	GAT	GTC	GAC	CTT	TGT	GAG	CCA	AGC	CCC	TGC	1456
Gly	Pro	Leu	Cys	Glu	Val	Asp	Val	Asp	Leu	Cys	Glu	Pro	Ser	Pro	Cys	
360				365					370						375	
CGG	AAC	GGC	GCT	CGC	TGC	TAT	AAC	CTG	GAG	GGT	GAC	TAT	TAC	TGC	GCC	1504
Arg	Asn	Gly	Ala	Arg	Cys	Tyr	Asn	Leu	Glu	Gly	Asp	Tyr	Tyr	Cys	Ala	
			380						385						390	
TGC	CCT	GAT	GAC	TTT	GGT	GGC	AAG	AAC	TGC	TCC	GTG	CCC	CGC	GAG	CCG	1552
Cys	Pro	Asp	Asp	Phe	Gly	Gly	Lys	Asn	Cys	Ser	Val	Pro	Arg	Glu	Pro	
			395					400						405		
TGC	CCT	GGC	GGG	GCC	TGC	AGA	GTG	ATC	GAT	GGC	TGC	GGG	TCA	GAC	GCG	1600
Cys	Pro	Gly	Gly	Ala	Cys	Arg	Val	Ile	Asp	Gly	Cys	Gly	Ser	Asp	Ala	
		410					415					420				
GGG	CCT	GGG	ATG	CCT	GGC	ACA	GCA	GCC	TCC	GGC	GTG	TGT	GGC	CCC	CAT	1648
Gly	Pro	Gly	Met	Pro	Gly	Thr	Ala	Ala	Ser	Gly	Val	Cys	Gly	Pro	His	
	425					430					435					
GGA	CGC	TGC	GTC	AGC	CAG	CCA	GGG	GGC	AAC	TTT	TCC	TGC	ATC	TGT	GAC	1696
Gly	Arg	Cys	Val	Ser	Gln	Pro	Gly	Gly	Asn	Phe	Ser	Cys	Ile	Cys	Asp	
440					445				450						455	
AGT	GGC	TTT	ACT	GGC	ACC	TAC	TGC	CAT	GAG	AAC	ATT	GAC	GAC	TGC	CTG	1744
Ser	Gly	Phe	Thr	Gly	Thr	Tyr	Cys	His	Glu	Asn	Ile	Asp	Asp	Cys	Leu	
			460						465					470		
GGC	CAG	CCC	TGC	CGC	AAT	GGG	GGC	ACA	TGC	ATC	GAT	GAG	GTG	GAC	GCC	1792
Gly	Gln	Pro	Cys	Arg	Asn	Gly	Gly	Thr	Cys	Ile	Asp	Glu	Val	Asp	Ala	
			475					480					485			
TTC	CGC	TGC	TTC	TGC	CCC	AGC	GGT	TGG	GAG	GGC	GAG	CTC	TGC	GAC	ACC	1840
Phe	Arg	Cys	Phe	Cys	Pro	Ser	Gly	Trp	Glu	Gly	Glu	Leu	Cys	Asp	Thr	
		490					495					500				
AAT	CCC	AAC	GAC	TGC	CTT	CCC	GAT	CCC	TGC	CAC	AGC	CGC	GGC	CGC	TGC	1888
Asn	Pro	Asn	Asp	Cys	Leu	Pro	Asp	Pro	Cys	His	Ser	Arg	Gly	Arg	Cys	
	505					510					515					
TAC	GAC	CTG	GTC	AAT	GAC	TTC	TAC	TGT	GCG	TGC	GAC	GAC	GGC	TGG	AAG	1936
Tyr	Asp	Leu	Val	Asn	Asp	Phe	Tyr	Cys	Ala	Cys	Asp	Asp	Gly	Trp	Lys	
520					525				530						535	
GGC	AAG	ACC	TGC	CAC	TCA	CGC	GAG	TTC	CAG	TGC	GAT	GCC	TAC	ACC	TGC	1984
Gly	Lys	Thr	Cys	His	Ser	Arg	Glu	Phe	Gln	Cys	Asp	Ala	Tyr	Thr	Cys	
			540						545					550		
AGC	AAC	GGT	GGC	ACC	TGC	TAC	GAC	AGC	GGC	GAC	ACC	TTC	CGC	TGC	GCC	2032
Ser	Asn	Gly	Gly	Thr	Cys	Tyr	Asp	Ser	Gly	Asp	Thr	Phe	Arg	Cys	Ala	
			555					560					565			
TGC	CCC	CCC	GGC	TGG	AAG	GGC	AGC	ACC	TGC	GCC	GTC	GCC	AAG	AAC	AGC	2080
Cys	Pro	Pro	Gly	Trp	Lys	Gly	Ser	Trp	Cys	Ala	Val	Ala	Lys	Asn	Ser	
		570					575					580				

**FIG. 2C**  
**SUBSTITUTE SHEET (RULE 26)**

11/20

AGC	TGC	CTG	CCC	AAC	CCC	TGT	GTG	AAT	GGT	GGC	ACC	TGC	GTG	GGC	AGC	2128
Ser	Cys	Leu	Pro	Asn	Pro	Cys	Val	Asn	Gly	Gly	Thr	Cys	Val	Gly	Ser	
585						590					595					
GGG	GCC	TCC	TTC	TCC	TGC	ATC	TGC	CGG	GAC	GGC	TGG	GAG	GGT	CGT	ACT	2176
Gly	Ala	Ser	Phe	Ser	Cys	Ile	Cys	Arg	Asp	Gly	Trp	Glu	Gly	Arg	Thr	
600					605					610					615	
TGC	ACT	CAC	AAT	ACC	AAC	GAC	TGC	AAC	CCT	CTG	CCT	TGC	TAC	AAT	GGT	2224
Cys	Thr	His	Asn	Thr	Asn	Asp	Cys	Asn	Pro	Leu	Pro	Cys	Tyr	Asn	Gly	
				620					625						630	
GGC	ATC	TGT	GTT	GAC	GGC	GTC	AAC	TGG	TTC	CGC	TGC	GAG	TGT	GCA	CCT	2272
Gly	Ile	Cys	Val	Asp	Gly	Val	Asn	Trp	Phe	Arg	Cys	Glu	Cys	Ala	Pro	
			635					640						645		
GGC	TTC	GCG	GGG	CCT	GAC	TGC	CGC	ATC	AAC	ATC	GAC	GAG	TGC	CAG	TCC	2320
Gly	Phe	Ala	Gly	Pro	Asp	Cys	Arg	Ile	Asn	Ile	Asp	Glu	Cys	Gln	Ser	
		650					655							660		
TCG	CCC	TGT	GCC	TAC	GGG	GCC	ACG	TGT	GTG	GAT	GAG	ATC	AAC	GGG	TAT	2368
Ser	Pro	Cys	Ala	Tyr	Gly	Ala	Thr	Cys	Val	Asp	Glu	Ile	Asn	Gly	Tyr	
		665				670					675					
CGC	TGT	AGC	TGC	CCA	CCC	GGC	CGA	GCC	GGC	CCC	CGG	TGC	CAG	GAA	GTG	2416
Arg	Cys	Ser	Cys	Pro	Pro	Gly	Arg	Ala	Gly	Pro	Arg	Cys	Gln	Glu	Val	
680					685					690					695	
ATC	GGG	TTC	GGG	AGA	TCC	TGC	TGG	TCC	CGG	GGC	ACT	CCG	TTC	CCA	CAC	2464
Ile	Gly	Phe	Gly	Arg	Ser	Cys	Trp	Ser	Arg	Gly	Thr	Pro	Phe	Pro	His	
				700					705					710		
GGA	AGC	TCC	TGG	GTG	GAA	GAC	TGC	AAC	AGC	TGC	CGC	TGC	CTG	GAT	GGC	2512
Gly	Ser	Ser	Trp	Val	Glu	Asp	Cys	Asn	Ser	Cys	Arg	Cys	Leu	Asp	Gly	
		715						720						725		
CGC	CGT	GAC	TGC	AGC	AAG	GTG	TGG	TGC	GGA	TGG	AAG	CCT	TGT	CTG	CTG	2560
Arg	Arg	Asp	Cys	Ser	Lys	Val	Trp	Cys	Gly	Trp	Lys	Pro	Cys	Leu	Leu	
		730					735							740		
GCC	GGC	CAG	CCC	GAG	GCC	CTG	AGC	GCC	CAG	TGC	CCA	CTG	GGG	CAA	AGG	2608
Ala	Gly	Gln	Pro	Glu	Ala	Leu	Ser	Ala	Gln	Cys	Pro	Leu	Gly	Gln	Arg	
		745				750								755		
TGC	CTG	GAG	AAG	GCC	CCA	GGC	CAG	TGT	CTG	CGA	CCA	CCC	TGT	GAG	GCC	2656
Cys	Leu	Glu	Lys	Ala	Pro	Gly	Gln	Cys	Leu	Arg	Pro	Pro	Cys	Glu	Ala	
		760				765				770					775	
TGG	GGG	GAG	TGC	GGC	GCA	GAA	GAG	CCA	CCG	AGC	ACC	CCC	TGC	CTG	CCA	2704
Trp	Gly	Glu	Cys	Gly	Ala	Glu	Glu	Pro	Pro	Ser	Thr	Pro	Cys	Leu	Pro	
				780					785					790		
CGC	TCC	GGC	CAC	CTG	GAC	AAT	AAC	TGT	GCC	CGC	CTC	ACC	TTG	CAT	TTC	2752
Arg	Ser	Gly	His	Leu	Asp	Asn	Asn	Cys	Ala	Arg	Leu	Thr	Leu	His	Phe	
				795				800						805		

FIG. 2D  
SUBSTITUTE SHEET (RULE 26)

12/20

AAC	CGT	GAC	CAC	GTG	CCC	CAG	GGC	ACC	ACG	GTG	GGC	GCC	ATT	TGC	TCC	2800
Asn	Arg	Asp	His	Val	Pro	Gln	Gly	Thr	Thr	Val	Gly	Ala	Ile	Cys	Ser	
	810						815					820				
GGG	ATC	CGC	TCC	CTG	CCA	GCC	ACA	AGG	GCT	GTG	GCA	CGG	GAC	CGC	CTG	2848
Gly	Ile	Arg	Ser	Leu	Pro	Ala	Thr	Arg	Ala	Val	Ala	Arg	Asp	Arg	Leu	
	825					830					835					
CTG	GTG	TTG	CTT	TGC	GAC	CGG	GCG	TCC	TCG	GGG	GCC	AGT	GCT	GTG	GAG	2896
Leu	Val	Leu	Leu	Cys	Asp	Arg	Ala	Ser	Ser	Gly	Ala	Ser	Ala	Val	Glu	
840					845					850					855	
GTG	GCC	GTG	TCC	TTC	AGC	CCT	GCC	AGG	GAC	CTG	CCT	GAC	AGC	AGC	CTG	2944
Val	Ala	Val	Ser	Phe	Ser	Pro	Ala	Arg	Asp	Leu	Pro	Asp	Ser	Ser	Leu	
			860						865					870		
ATC	CAG	GGC	GCG	GCC	CAC	GCC	ATC	GTG	GCC	GCC	ATC	ACC	CAG	CGG	GGG	2992
Ile	Gln	Gly	Ala	Ala	His	Ala	Ile	Val	Ala	Ala	Ile	Thr	Gln	Arg	Gly	
	875						880						885			
AAC	AGC	TCA	CTG	CTC	CTG	GCT	GTC	ACC	GAG	GTC	AAG	GTG	GAG	ACG	GTT	3040
Asn	Ser	Ser	Leu	Leu	Leu	Ala	Val	Thr	Glu	Val	Lys	Val	Glu	Thr	Val	
	890						895				900					
GTT	ACG	GGC	GGC	TCT	TCC	ACA	GGT	CTG	CTG	GTG	CCT	GTG	CTG	TGT	GGT	3088
Val	Thr	Gly	Gly	Ser	Ser	Thr	Gly	Leu	Leu	Val	Pro	Val	Leu	Cys	Gly	
	905					910					915					
GCC	TTC	AGC	GTG	CTG	TGG	CTG	GCG	TGC	GTG	GTC	CTG	TGC	GTG	TGG	TGG	3136
Ala	Phe	Ser	Val	Leu	Trp	Leu	Ala	Cys	Val	Val	Leu	Cys	Val	Trp	Trp	
920					925					930					935	
ACA	CGC	AAG	CGC	AGG	AAA	GAG	CGG	GAG	AGG	AGC	CGG	CTG	CCG	CGG	GAG	3184
Thr	Arg	Lys	Arg	Arg	Lys	Glu	Arg	Glu	Arg	Ser	Arg	Leu	Pro	Arg	Glu	
				940					945					950		
GAG	AGC	GCC	AAC	AAC	CAG	TGG	GCC	CCG	CTC	AAC	CCC	ATC	CGC	AAC	CCC	3232
Glu	Ser	Ala	Asn	Asn	Gln	Trp	Ala	Pro	Leu	Asn	Pro	Ile	Arg	Asn	Pro	
			955				960						965			
ATT	GAG	CGG	CCG	GGG	GGG	CAC	AAG	GAC	GTG	CTC	TAC	CAG	TGC	AAG	AAC	3280
Ile	Glu	Arg	Pro	Gly	Gly	His	Lys	Asp	Val	Leu	Tyr	Gln	Cys	Lys	Asn	
	970						975					980				
TTC	ACT	CCA	CCG	CCG	CGC	AGG	CGC	TGC	CCG	GGC	CGG	CCG	GCC	ACG	CGG	3328
Phe	Thr	Pro	Pro	Pro	Arg	Arg	Arg	Cys	Pro	Gly	Arg	Pro	Ala	Thr	Arg	
	985					990					995					
CCG	TCA	GGG	AGG	ATG	AGG	AGG	ACG	AGG	ATC	TTG	GCC	GCG	GTG	AGG	AGG	3376
Pro	Ser	Gly	Arg	Met	Arg	Arg	Thr	Arg	Ile	Leu	Ala	Ala	Val	Arg	Arg	
1000					1005				1010					1015		
ACT	CCC	TGG	AGG	CGG	AGA	AGT	TCC	TCT	CAC	ACA	AAT	TCA	CCA	AAG	ATC	3424
Thr	Pro	Trp	Arg	Arg	Arg	Ser	Ser	Ser	His	Thr	Asn	Ser	Pro	Lys	Ile	
				1020					1025					1030		

**FIG. 2E**  
**SUBSTITUTE SHEET (RULE 26)**

13/20

CTG GCC GCT CGC CGG GGA GGC CGG CCC ACT GGG CCT CAG GCC CCA AAG 3472  
 Leu Ala Ala Arg Arg Gly Gly Arg Pro Thr Gly Pro Gln Ala Pro Lys  
 1035 1040 1045  
 TGG ACA ACC GCG CGG TCA GGA GCA TCA ATG AGG CCC GCT ACG TCG GCA 3520  
 Trp Thr Thr Ala Arg Ser Gly Ala Ser Met Arg Pro Ala Thr Ser Ala  
 1050 1055 1060  
 AGG GAA GTA GGG CGG CTG CAG CTG GGC CGG GAC CCA GGG CCC TCG GTG 3568  
 Arg Glu Val Gly Arg Leu Gln Leu Gly Arg Asp Pro Gly Pro Ser Val  
 1065 1070 1075  
 GGA GCC ATG CCG TCT GCC GGA CCC GGA GGC CGA GGC CAT GTG CAT AGT 3616  
 Gly Ala Met Pro Ser Ala Gly Pro Gly Gly Arg Gly His Val His Ser  
 1080 1085 1090 1095  
 TTC TTT ATT TTG TGT AAA AAA ACC ACC AAA AAC AAA AAC CAA ATG TTT 3664  
 Phe Phe Ile Leu Cys Lys Lys Thr Thr Lys Asn Lys Asn Gln Met Phe  
 1100 1105 1110  
 ATT TTC TAC GTT TCT TTA ACC TTG TAT AAA TTA TTC AGT AAC TGT CAG 3712  
 Ile Phe Tyr Val Ser Leu Thr Leu Tyr Lys Leu Phe Ser Asn Cys Gln  
 1115 1120 1125  
 GCT GAA AAC AAT GGA GTA TTC TCG GAT AGT TGC TAT TTT TGT AAA GTA 3760  
 Ala Glu Asn Asn Gly Val Phe Ser Asp Ser Cys Tyr Phe Cys Lys Val  
 1130 1135 1140  
 GCC GTG CGT GGC ACT CGC TGT ATG AAA GGA GAG AGC AAA GGG TGT CTG 3808  
 Ala Val Arg Gly Thr Arg Cys Met Lys Gly Glu Ser Lys Gly Cys Leu  
 1145 1150 1155  
 CGT CGT CAC CAA ATC GTC GCG TTT GTT ACC AGA GGT TGT GCA CTG TTT 3856  
 Arg Arg His Gln Ile Val Ala Phe Val Thr Arg Gly Cys Ala Leu Phe  
 1160 1165 1170 1175  
 ACA GAA TCT TCC TTT TAT TCC TCA CTC GGG TTT CTC TGT GCT CCA GGC 3904  
 Thr Glu Ser Ser Phe Tyr Ser Ser Leu Gly Phe Leu Cys Ala Pro Gly  
 1180 1185 1190  
 CAA AGT GCC GGT GAG ACC CAT GGC TGT GTT GGT GTG GCC CAT GGC TGT 3952  
 Gln Ser Ala Gly Glu Thr His Gly Cys Val Gly Val Ala His Gly Cys  
 1195 1200 1205  
 TGG TGG GAC CCG TGG CTG ATG GTG TGG CCT GTG GCT GTC GGT GGG ACT 4000  
 Trp Trp Asp Pro Trp Leu Met Val Trp Pro Val Ala Val Gly Gly Thr  
 1210 1215 1220  
 CGT GGC TGT CAA TGG GAC CTG TGG CTG TCG GTG GGA CCT ACG GTG GTC 4048  
 Arg Gly Cys Gln Trp Asp Leu Trp Leu Ser Val Gly Pro Thr Val Val  
 1225 1230 1235

FIG. 2F

SUBSTITUTE SHEET (RULE 26)

14/20

GGT GGG ACC CTG GTT ATT GAT GTG GCC CTG GCT GCC GGC ACG GCC CGT 4096  
Gly Gly Thr Leu Val Ile Asp Val Ala Leu Ala Ala Gly Thr Ala Arg  
1240 1245 1250 1255  
GGC TGT TG ACGCACCTGT GGTGTGTTAGT GGGGCCTGAG GTCATCGGCG TGGCCCAAGG 4154  
Gly Cys  
CCGGCAGGTC AACCTCGCGC TTGCTGGCCA GTCCACCCTG CCTGCCGTCT GTGCTTCCTC 4214  
CTGCCCAGAA CGCCCGCTCC AGCGATCTCT CCACTGTGCT TTCAGAAGTG CCCTTCCTGC 4274  
TGCGCAGTTC TCCCATCCTG GGACGGCGGC AGTATTGAAG CTCGTGACAA GTGCCTTCAC 4334  
ACAGACCCCT CGCAACTGTC CACGCGTGCC GTGGCACCAG GCGCTGCCCA CCTGCCGGCC 4394  
CCGGCCGCCC CTCCTCGTGA AAGTGCATTT TTGTAAATGT GTACATATTA AAGGAAGCAC 4454  
TCTGTATAAA AAAAAAAAAAC CGGAATTCC 4483

**FIG. 2G****SUBSTITUTE SHEET (RULE 26)**

15/20

CAGGTGGCGTCAGCATCGGGACAGTTCGAGCTGGAGATCTTATCCGTGCAGAATGTGAACGGCGTGCT  
GCAGAACGGGAAGTCTGCGACGGCACTCGAAACCCCGGAGATAAAAAGTGCACCAGAGATGAGTGTG  
ACACCTACTTTAAAGTTTGCCTGAAGGAGTACCAGTCGCGGGTCACTGCTGGCGGCCCTTGACGCTTC  
GGATCCAAATCCACCCCTGTATCGGCGGGAATACCTTCAATTTAAAGTACAGCCGGAATAATGAAAA  
GAACCGGATTGTTATCCCTTTACGTTGCGCTGGCCGAGATCCTACACGTTGCTTGTGAGGCATGGG  
ATTACAATGATAACTCTACTAATCCCGATCGCATAATTGAGAAGGCATCCCACTCTGGCATGATCAAT  
CCAAGCCGTCACTGGCAGACGTTGAAACATAACACAGGAGCTGCCCACTTTGAGTATCAAATCCGTGT  
GACTTGCGCAGAACATTACTATGGCTTTGGATGCAACAAGTTTTGTCGACCGAGAGATGACTTCTTCA  
CTCACCATACTGTGACCAGAATGGCAACAAAACCTGCTTGAAGGCTGGACGGGACCAGAATGCAAC  
AAAGCTATTTGTCGTGAGGGATGTAGCCCCAAGCATGGTTCTTGACAGTTCAGGAGAGTGCAGGTG  
TCAGTATGGATGGCAAGGCCAGTACTGTGATAAGTGCATTCCACACCCGGGATGTGTCCATGGCACTT  
GCATTGAACCATGGCAGTGCCTCTGTGAAACCAACTGGGGTGGTCACTCTGTGACAAAGACCTGAAC  
TACTGTGGAACCCACCCACCTGTTTGAATGGTGGTACCTGCAGCAACACTGGCCCCGATAAATACCA  
GTGTTCTGCCCTGAGGGTTACTCAGGACAGAAGTGTGAAATAGCGGAGCATGCGTGCCTCTCTGATC  
CGTGCCACAACGGAGGAAGCTGCCTAGAAACGTCTACAGGATTTGAATGTGTGTGTGCACCTGGCTGG  
GCTGGACCAACTTGCACTGATAATATTGATGATTGTTCTCCAAATCCCTGTGGTCACTGGAGGAAGTGT  
CCAAGATCTAGTTGATGGATTAAAGTGTATTTGCCACCTCAGTGGACTGGCAAACATGCCAGCTAG  
ATGCGAATGAATGTGAGGGCAAACCTGTGTCAATGCCAACTCCTGCAGGAACCTGATTGGCAGCTAC  
TATTGTGACTGCATTACTGGCTGGTCTGGCCACAAGTGTGATATAAATATTAATGATTGTGCTGGACA  
ATGTGAGAATGGAGGATCCTGTCGGGACTTGGTTAATGGTTATCGGTGCATCTGTTACCTGGCTATG  
CAGGAGATCACTGTGAGAAAGACATCAATGAATGTGCAAGTAACCCTTGATGAATGGGGGTCACTGC  
CAGGATGAAATCAATGGATTCCAATGTCTGTGTCCTGCTGGTTTCTCAGGAAACCTCTGTGAGCTGGA  
TATAGACTACTGTGAGCCAAACCTTGCCAGAACGGTGCCCACTGCTTCAATCTTGCTATGGACTATT  
TCTGTAAGTGCCTGAAGATTACGAAGGCAAGAACTGCTCCACCTGAAAGATCACTGCCGCACAAGT  
CCTTGTGAAGTAATCGACAGCTGTACAGTGGCAGTGGCTTCTAACAGCACACCAGAAGGAGTTCGTTA  
CATTTCTTCAAATGTCTGTGGTCTCATGGAAAATGCAAGAGCCAAGCAGGTGGAAAATTCACCTGTG  
AATGCAACAAAGGATTCACTGGCACCTACTGTGATGAGAATATCAATGACTGTGAGAGCAACCCCTGT  
AAAAATGGTGGCACTTGTATTGACGGTGTAACTCCTACAAATGTATTTGTAGTGATGGATGGGAAGG  
AACATATTGTGAAACAAATATTAATGACTGCAGTAAAAACCCCTGCCACAATGGAGGAAGTTCGCCGAG  
ACTTGGTCAATGACTTCTTCTGTGAATGTAAAAATGGGTGGAAAGGAAAAACTTGCCACTCTCGTGAC  
AGCCAGTGTGATGAGGCAACATGCAATAATGGAGGAACATGTTATGATGAGGGGGACACTTCAAGTG  
CATGTGTCCTGCAGGATGGGAAGGAGCCACTTGTAAATATAGCAAGGAACAGCAGCTGCCTGCCAAACC  
CCTGTCACAATGGTGGTACCTGTGTAGTTAGTGGGGATTCTTCACTTGTGTCTGCAAGGAGGGCTGG  
GAAGGACCGACATGTACTCAGAACACAAATGACTGCAGTCTCATCTTGTACAAACAGTGGTACTTG  
TGTGGATGGAGACAAGTGGTACCGCTGTGAGTGCCTCCCGGCTTCGCAGGTCCCGACTGTAGGATCA  
ACATCAATGAATGTCAGTCTTACCCTGTGCCTTTGGGGCTACTTGTGTGGATGAAATTAATGGGTAC  
CGTTGCATTTGTCCACCGGGTGCAGTGGTCCAGGATGCCAGGAAGTTACAGGGAGGCCTTGCTTTAC  
CAGTATTCGAGTAATGCCAGACGGTGTAAAGTGGGATGATGACTGTAATACTTGTGAGTGTGTTGAATG  
GAAAAGTCACCTGTTCTAAGGTTTGGTGTGGTCTCGACCTTGATAATACATGCCAAAGGTCATAAT  
GAATGCCAGCTGGACACGCTTGTGTTCTGTAAAGAAGACCATTGTTTCACTCATCCTTGTGCTGC

FIG. 3A

SUBSTITUTE SHEET (RULE 26)

16/20

AGTGGGTGAATGCTGGCCTTCTAATCAGCAGCCTGTGAAGACCAAATGCAATTCTGATTCTTATTACC  
AAGATAATTGTGCCAACATCACCTTCACCTTTAATAAGGAAATGATGGCACCAGGCCTTACCACGGAG  
CACATTTGCAGTGAATTGAGGAATCTGAATATCCTGAAGAATGTTTCTGCTGAATATTCCATCTATAT  
TACCTGTGAGCCTTCACACTTGGCAAATAATGAAATACATGTTGCTATTTCTGCTGAAGATATAGGAG  
AAGATGAAAACCCAATCAAGGAAATCACAGATAAGATTATTGACCTTGTGAGTAAGCGTGATGGAAAC  
AACACACTAATTGCTGCAGTCGCAGAAGTCAGAGTACAAAGGCGACCAGTTAAGAACAAAACAGATTT  
CTTGGTGCCATTACTGAGCTCAGTCTTAACAGTAGCCTGGATCTGCTGTCTGGTAACTGTTTTCTATT  
GGTGCAATTCAAAAGCGCAGAAAGCAGAGCAGCCATACTCACACAGCATCTGATGACAACACCACCAAC  
AACGTAAGGGAGCAGCTGAATCAGATTA AAAACCCCATAGAGAAACACGGAGCAAATACTGTTCCAAT  
TAAAGACTATGAAAACAAAACTCTAAAATCGCCAAAATAAGGACGCACAATTCAGAAGTGGAGGAAG  
ATGACATGGACAAACACCAGCAAAGGCCCGGTTTGCCAAGCAGCCAGCGTACACTTTGGTAGACAGA  
GATGAAAAGCCACCCAACAGCACACCCACAAAACACCCAACTGGACAAATAAACAGGACAACAGAGA  
CTTGAAAAGTGCACAAAGTTTAAATAGAATGGAGTACATTGTATAG

**FIG. 3B****SUBSTITUTE SHEET (RULE 26)**

17/20

QVASASGOFE LEILSVQNVN GVLQNGNCCD GTRNPGDKKC TRDECDTYFK 50  
 VCLKEYQSRV TAGGPCSFGS KSTPVIGGNT FNLKYSRNNE KNRIVIPFSF 100  
 AWPRSYTLLV EAWDYNDNST NPDRIIEKAS HSGMINPSRQ WQTLKHNTGA 150  
 AHFEYQIRVT CAEHYYGFGC NKFCRPRDDF FTEHTCDQNG NKTCLGWTG 200  
 \*\*\*\*\*DSL DOMAIN\*\*\*\*\*  
 PECNKAICRQ GCSPKHGSCT VPGEERCQYG WQGQYCDKCI PHPGCVHGTG 250  
 \*\*\* <-----EGF 1----->-----  
 IEPWQCLCET NWGGQLCDKD LNYCGTHPPC LGGGTCSTNG PDKYQCSCPE 300  
 -----EGF 2----->-----EGF 3-----  
 GYSGQNCIEA EHACLSDPCH NGGSCLETST GFECVCAPGW AGPTCTDNID 350  
 ----->-----EGF 4-----  
 DCSPNPCGHG GTCQDLVDGF KCICPPQWTG KTCQLDANEC EGKPCVNANS 400  
 >-----EGF 5----->-----  
 CRNLIGSYYC DCITGWSGHN CDININDCRG QCQNGGSCRD LVNGYRCICS 450  
 -----EGF 6----->-----EGF 7---  
 PGYAGDHCEK DINECASNPC MNGGHCODEI NGFQCLCPAG FSGNLCQLDI 500  
 ----->-----EGF 8-----  
 DYCEPNPCQN GAQCFNLAMD YFCNCPEDYE GKNCSHLKDH CRTTPCEVID 550  
 ->-----EGF 9----->-----  
 SCTVAVASNS TPEGVRYISS NVCGPHGKCK SQAGGKFTCE CNKGFTGTYC 600  
 -----EGF 10-----  
 HENINDCESN PCKNGGTCID GVNSYKICIS DGWEGTYCET NINDCSKNPC 650  
 ----->-----EGF 11----->-----  
 HNGGTCRDLV NDFICECKNG WKGKTCHSRD SQCDEATCNN GGTCYDEGDT 700  
 -----EGF 12----->-----  
 FKCMCPAGWE GATCNIARNS SCLPNPCHNG GTCVVS GDSF TCVCKEGWEG 750  
 EGF 13----->-----EGF 14-----  
 PTCTQNTNDC SPHPCYNSGT CVDGDNWYRC ECAPGFAGPD CRININECQS 800  
 ----->-----EGF 15----->-----  
 SPCAFGATCV DEINGYRCIC PPGRSGPGCQ EVTGRPCFTS IRVMPDGAKW 850  
 -----EGF 16----->-----  
 DDDCNTCQCL NGKVTCSKVW CGPRPCIIHA KGHNECPAGH ACVPVKEDHC 900  
 <- Cysteine-Rich Region  
 FTHPCAAGE CWPSNQPVK TKCNSDSYYQ DNCANITFTF NKEMMAPGLT 950  
 ->  
 TEHICSELRN LNILKNVSAE YSIYITCEPS HLANNEIHVA ISAEDIGEDE 1000

FIG. 4A

SUBSTITUTE SHEET (RULE 26)



18/20

NPIKEITDKI IDLVSKRDGN NTLIAAAEV RVQRRPVKNK TDFLVPLLSS 1050  
VLTVAWICCL VTVFYWCIOK RRKQSSHTHT ASDDNTTNNV REQLNQIKNP 1100  
IEKHGANTVP IKDYENKNSK IAKIRTHNSE VEEDDMDKHQ QKARFAKQPA 1150  
YTLVDRDEKP PNSTPTKHPN WTNKQDNRDL ESAQSLNRME YIV 1193

**FIG. 4B****SUBSTITUTE SHEET (RULE 26)**

19/20

DmDelta	SGSFELRLKY	FSNDHGRDNE	GRCCS-GESD	GATGKCL-GS	OKTHFRRLCLK	48
CSer	SGQFELEILS	VQNVNGVLQN	GNCCD-GTRN	PGDKKCTRDE	QDTYFKVCLK	49
DmSer	AGNFELEILE	ISNTNSHLLN	GYCOGMPAEL	RATKTIGCSP	QTTAFRLCLK	50
DmDelta	HYQATIDTTS	QCTYGDVITP	ILGENSVNLT	DAQRFQNKGF	TNPIDPFPSF	98
CSer	EYQSRVTAGG	PCSFGSKSTP	VIGGNTFNL-	—KYSRNE	KNRIVIPFSF	95
DmSer	EYQTEQGAS	ISTGCSFGNA	TTKILGGSS-	—FVLSDPG	VGAIVLPFTIF	96
DmDelta	SWPGTFS	IV EAWHDTNNSG	NARTNKLL	TO RLLVQOVLEV	SSEWNTNKSE	148
CSer	AWPRSYTL	LV EAWDYNDNS-	-TNPDR-II	E KASHSCMINP	SROWTLKHN	142
DmSer	RMTKSFTL	IL QALDMYNTS-	YPDAER-LI	E ETSYSGVILP	SPEWTL DHI	144
DmDelta	SQYTSLEYDF	RVITQDLNYYG	SGCAKFCRPR	DDSFCHSTCS	ETGEIICLTG	198
CSer	TGAAHFEYDI	RVITQAEHYG	FGONKFCRPR	DDFFTHHTCD	QNGNKTCLCG	192
DmSer	GRNARITYRV	RVQCAVTYYN	TTCTTCRPR	DDQEGHYACC	SEGQKLCING	194
===== D S L DOMAIN =====						
DmDelta	WGGDYC					204
CSer	WTGPEC					198
DmSer	WGGVNC					200

FIG.5

20/20

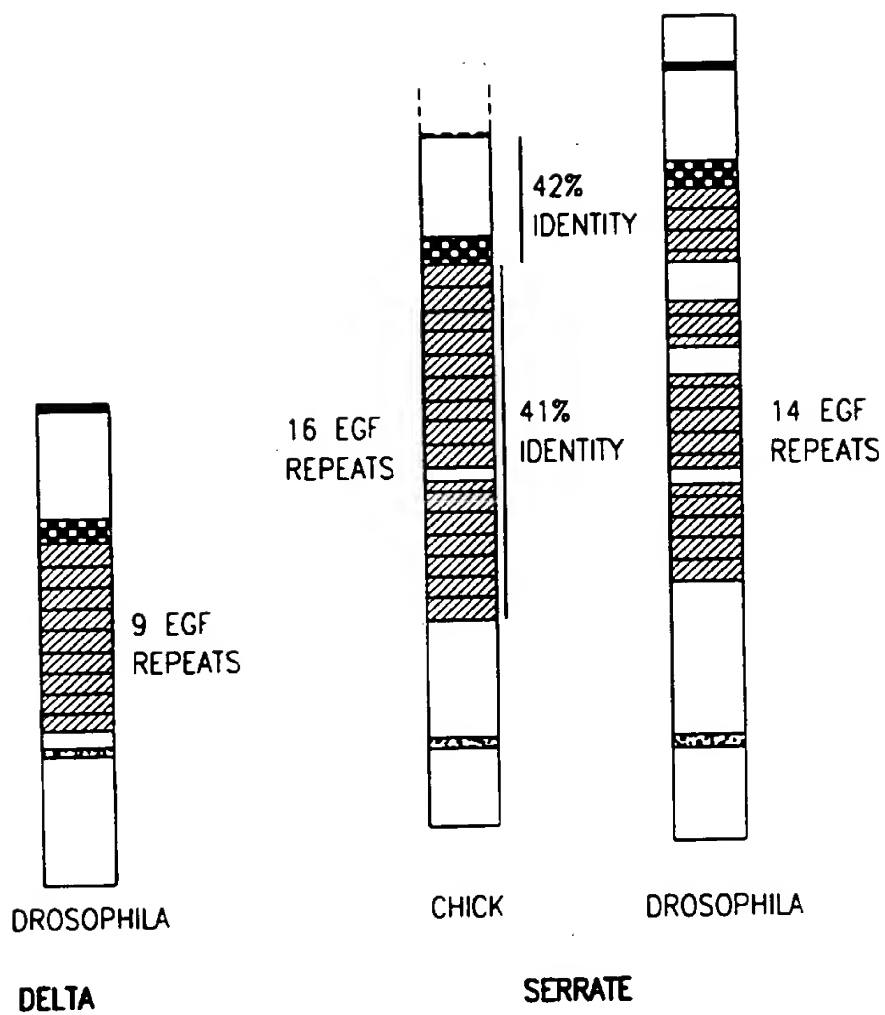


FIG.6

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US96/03172

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : C07K14/00; A01N 37/18, 43/04; C07H 17/00; C12N 5/00; C12P 21/06  
US CL : 530/350, 387.1; 514/2, 44/536/23.1/ 435/69.1, 240.1  
According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 530/350, 387.1; 514/2, 44/536/23.1/ 435/69.1, 240.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched  
None

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)  
APS, Dialog

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, A	GRAY et al. Human ligands of the notch receptor. Society for Neuroscience Abstracts, 11 November 1995, Vol. 21, No. 1-3, page 1524, abstract 601.1. See entire abstract.	1-82
P, A	CHITNIS et al. Primary neurogenesis in Xenopus embryos regulated by a homologue of the Drosophila neurogenic gene Delta. Nature, 29 June 1995, Vol. 375, pages 761-766. See entire document.	1-82
P, A	HENRIQUE et al. Expression of a Delta homologue in prospective neurons in the chick. Nature, 29 June 1995, Vol. 375, pages 787-790. See entire document.	1-82



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:	* T	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
* A* document defining the general state of the art which is not considered to be of particular relevance	* X*	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
* E* earlier document published on or after the international filing date	* Y*	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
* I* document which may throw doubt on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	* G*	document member of the same patent family
* O* document referring to an oral disclosure, use, exhibition or other means		
* P* document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 08 JULY 1996	Date of mailing of the international search report 30 JUL 1996
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-2230	Authorized officer Karen Cochran Carlson, Ph.D. Telephone No. (703) 308-0196

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US96/03172

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐  
☐

- The additional search fees were accompanied by the applicant's protest.  
No protest accompanied the payment of additional search fees.

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US96/03172

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, A	LINDSELL et al. Jagged: A mammalian ligand that activates Notch1. Cell. 24 March 1995, Vol. 80, pages 909-917. See entire document.	1-82

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US96/03172

## BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING

This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, claim(s) 1-29, 34-68, 78, and 79, drawn to DNA and amino acid sequences encoding Serrate protein.

Group II, claim(s) 30-33, drawn to an antibody against Serrate protein.

Group III, claim(s) 69-73 and 77, drawn to a method of treating disease with Serrate protein.

Group IV, claim(s) 74 and 75, drawn to a method of treating disease with DNA encoding Serrate protein - gene therapy.

Group V, claim(s) 76, drawn to a method of treating a disease with the antibody against Serrate protein.

Group VI, claim(s) 80, drawn to a method of inhibiting expression of Serrate protein using antisense DNA.

Group VII, claim(s) 81, drawn to method for diagnosing a disease via notch:Serrate protein binding assay.

Group VIII, claim(s) 82, drawn to a method for diagnosing a disease via measuring Serrate protein levels.

The inventions listed as Groups I through VIII do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features, for the following reasons:

Groups I and Group II are structurally different compounds/compositions having unique properties not shared by the other. A reference anticipating or rendering obvious the compound/composition of Group I would not necessarily anticipate or make obvious the compound/composition of Group II. Groups III-VIII are properly grouped separately from the main invention of Group I, pursuant to 37 CFR 1.475(d). Therefore, the groupings lack the same or corresponding special technical feature.

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US96/15651

## BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING

This ISA found multiple inventions as follows:

Group I, claims 1-33, 35, 39, 41-43, drawn to a method for expansion of precursor cells.

Group II, claim 34, drawn to a method of therapy.

Group III, claims 36-38, drawn to a method of inhibiting cell growth.

The inventions listed as Groups I, II, and III do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Precursor cells may be expanded by different methods, such as the addition of growth factors; an enormous variety of therapeutics are available for a wide range of pathological conditions; cell growth may be inhibited by a variety of methods including temperature control and addition of inhibitory agents.